Robust and efficient estimation of the residual scale in linear regression

Stefan Van Aelst[a,1,*], Gert Willems[a], Ruben H. Zamar[b,2]

[a]*Dept. of Applied Mathematics and Computer Science, Ghent University Krijgslaan 281 S9, B-9000 Gent, Belgium*
[b]*Dept. of Statistics, University of British Columbia 6356 Agricultural Road, Vancouver, BC, Canada V6T-1Z2*

Abstract

Robustness and efficiency of the residual scale estimators in the regression model is important for robust inference. We introduce the class of robust *generalized M-scale estimators* for the regression model, derive their influence function and gross-error sensitivity, and study their maxbias behavior. In particular, we find overall minimax bias estimates for the general class and also for well-known subclasses. We pose and solve a Hampel's-like optimality problem: we find generalized M-scale estimators with maximal efficiency subject to a lower bound on the *global* and *local* robustness of the estimators.

Keywords: robust scale, maxbias, influence function, gross-error sensitivity, efficiency
2010 MSC: 62F35,62J05

1. Introduction

In linear regression, the classical estimators for the regression coefficients and error scale are the well-known least squares estimators. These estimators are optimal under normal errors but extremely sensitive to outliers. This is particularly the case for the residual scale estimator.

Much attention has been paid in the statistical literature to robust and efficient estimation of the regression parameters. In this context, robust residual scale estimators are sometimes proposed as well, but the focus remains on the regression parameters. See e.g. Martin, Yohai, and Zamar (1989); He and Simpson (1993); Berrendero, Mendes, and Tyler (2007). For the location-scale

*Corresponding author
Email addresses: Stefan.VanAelst@UGent.be (Stefan Van Aelst), Gert11Willems@gmail.com (Gert Willems), ruben@stat.ubc.ca (Ruben H. Zamar)

[1]Research supported by a grant of the Fund for Scientific Research-Flanders (FWO-Vlaanderen) and by IAP research network grant nr. P6/03 of the Belgian government (Belgian Science Policy)
[2]Research supported by NSERC

model, some attention has been given to the estimation of the scale parameter. See for example Iglewicz (1982); Martin et al. (1989); Martin and Zamar (1993); Rousseeuw and Croux (1993); Randal (2008) and Croux and Haesbroeck (2001).

Robust residual scale estimates play an important role in robust inference for the regression model such as the construction of confidence/prediction intervals, testing of hypotheses and model selection. The properties of such robust inference procedures heavily depend on the parameter estimators involved (see e.g. Heritier and Ronchetti, 1994; Van Aelst and Willems, 2011). If the scale is involved in the inference, it is thus desirable to use a highly robust and highly efficient scale estimator to obtain a reliable and effective inference procedure.

Therefore, we study the statistical properties – robustness and efficiency – of a large class of robust residual scale estimators, which we call *generalized M-scales*. This class includes M-scales (Huber, 1964), S-scales (Rousseeuw and Yohai, 1984), and τ-scales (Yohai and Zamar, 1988) as particular cases. We show that the influence function (IF) and breakdown point (BDP) properties do not suffice to characterize the robustness behavior of generalized M-scales because many generalized M-scales can be constructed with the same IF and BDP that still exhibit quite different robustness performance. Therefore, we study the maxbias of generalized M-scales which is a more overall measure of the robustness of an estimator We investigate the maxbias behavior of generalized M-scales and determine which regression estimators must be used to maximize the robustness of the resulting scale estimator. Moreover, we find scale estimators that maximize efficiency under the central model subject to a bound on the gross-error sensitivity and breakdown point.

We now introduce some definitions and notation used throughout this paper. Consider n observations $(y_i, \mathbf{z}_i^t)^t \in \mathbb{R}^p$ and the linear regression model

$$y_i = \mathbf{x}_i^t \boldsymbol{\beta}_0 + \sigma_0 u_i, \qquad 1, \ldots, n, \tag{1}$$

where $\mathbf{x}_i = (1, \mathbf{z}_i^t)^t$. Under this central model, the errors u_i are assumed to be independent and identically distributed with a common distribution F_0, which is symmetric around zero and has scale one. Moreover, the errors are assumed to be independent of the predictors $\mathbf{z}_i$. We consider regression models with random predictors which is the standard when studying maxbias properties of regression estimators and allows us to obtain overall results that do not depend on a particular design. The distribution of the predictors under the central model is denoted by G_0 and the common joint distribution of $(y_i, \mathbf{z}_i^t)^t$ is denoted by H_0. To allow for a fraction $0 \leq \epsilon \leq \frac{1}{2}$ of outliers we assume that the actual underlying distribution H_ϵ of $(y_i, \mathbf{z}_i^t)^t; i = 1, \ldots, n$ belongs to the contamination neighborhood

$$\mathcal{H}_\epsilon = \{H_\epsilon : \ H_\epsilon = (1-\epsilon)H_0 + \epsilon H^*\},$$

where H^* is an arbitrary and unspecified distribution. Robust and efficient estimators for $\boldsymbol{\beta}_0$ or σ_0 are expected to perform relatively well for any $H_\epsilon \in \mathcal{H}_\epsilon$ with $0 \leq \epsilon < \epsilon_n^*$, for some ϵ_n^* which would preferably be close or equal to $\frac{1}{2}$. Efficiency of the estimators is measured by their performance at the central model $H = H_0$ (i.e. $\epsilon = 0$).

For any $\mathbf{v} \in \mathbb{R}^p$ let $r_i(\mathbf{v}) = y_i - \mathbf{x}_i^t \mathbf{v}$ denote the corresponding residuals. The most common estimators of the error scale parameter σ_0 are based on the residuals $r_i(\hat{\boldsymbol{\beta}}_n)$ for some regression estimator $\hat{\boldsymbol{\beta}}_n$. We consider the following general class of residual scale estimators that we call *generalized M-scales*:

$$\hat{\sigma}_n(\hat{s}_n, \hat{\boldsymbol{\beta}}_n) = \hat{s}_n \sqrt{\frac{1}{b\,n} \sum_{i=1}^{n} \rho\left(\frac{r_i(\hat{\boldsymbol{\beta}}_n)}{\hat{s}_n}\right)}. \tag{2}$$

Here $\hat{s}_n$ is an arbitrary *initial scale estimator*, perhaps based on residuals $r_i(\tilde{\boldsymbol{\beta}}_n)$ corresponding to a given regression fit $\tilde{\boldsymbol{\beta}}_n$. The constant b is set equal to $E_{F_0}\{\rho(u)\}$ to obtain a consistent scale estimator at the central model. Often the error distribution F_0 is assumed to be the standard Gaussian distribution to obtain a consistent scale estimator at this model, provided that the initial scale estimator $\hat{s}_n$ is also consistent.

The generalized M-scale estimators considered in this paper are based on loss functions $\rho : \mathbb{R} \to \mathbb{R}$ that satisfy the following conditions:

(A1) ρ is symmetric, bounded and nondecreasing on $[0, \infty]$ with $\rho(0) = 0$. Moreover, ρ is differentiable at all $x \in \mathbb{R}$, except perhaps at a finite number of points.

(A2) The error distribution $F_0(x)$ has a density $f_0(x)$, which is symmetric, continuous, and strictly decreasing for $u \geq 0$.

Some results will require an additional assumption:

(A3) $g\,(s) = E_{F_0}\left(s^2 \rho(\frac{u}{s})\right)$ is non-decreasing for all $s \geq 0$.

Note that (A3) holds if, for instance, $2\rho(u) - \rho'(u)u \geq 0$, which is true for several well known ρ functions including Tukey's biweight loss function. The symmetry of the error distribution in Assumption (A2) is natural and commonly made in robust regression. Moreover, it is needed to prove our results. Assumption (A1) implies that we can assume without loss of generality that $\lim_{t\to\infty} \rho(t) = 1$ when convenient. The derivative of the loss function $\rho(x)$, which exists according to Assumption (A1), is denoted by $\psi(x)$.

One would expect that a robust choice for the initial estimators $\hat{s}_n$ and $\hat{\boldsymbol{\beta}}_n$ combined with an efficient choice for the loss function ρ would lead to a highly robust and efficient estimator $\hat{\sigma}_n(\hat{s}_n, \hat{\boldsymbol{\beta}}_n)$. In fact, we will show that our class of generalized M-scales includes estimators that can achieve high efficiency at H_0 without compromising their robustness.

To derive the asymptotic properties of the scale estimators, we introduce the generalized M-scale functional corresponding to the generalized M-scale estimator in (2). For any distribution H on $\mathbb{R}^p$ the generalized M-scale functional is defined as

$$\hat{\sigma}(H; \hat{s}, \hat{\boldsymbol{\beta}}) = \hat{s}(H) \sqrt{\frac{1}{b}\, E_H \rho\left(\frac{y - \mathbf{x}^t \hat{\boldsymbol{\beta}}(H)}{\hat{s}(H)}\right)}, \tag{3}$$

where $\hat{s} = \hat{s}(H)$ and $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(H)$ are scale and regression functionals corresponding to the estimators $\hat{s}_n$ and $\hat{\boldsymbol{\beta}}_n$, respectively.

We now give some examples of special subclasses of generalized M-scale estimators defined by (2)-(3).

Example 1 (M-scales). Given a preliminary regression estimator $\tilde{\boldsymbol{\beta}}_n$, the corresponding M-scale estimator $\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)$ is implicitly defined as a solution, in s, to the equation

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i(\tilde{\boldsymbol{\beta}}_n)}{s}\right) = b, \tag{4}$$

see Huber (1964, 1981). Note that M-scales are indeed an example of generalized M-estimators. In fact, if in (2) we set $\hat{s}_n = \hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)$ and take the regression estimator $\hat{\boldsymbol{\beta}}_n$ equal to the preliminary regression estimator $\tilde{\boldsymbol{\beta}}_n$, then it immediately follows from (4) that the generalized M-scale in (2) reduces to $\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)$, that is $\hat{\sigma}_n(\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n), \tilde{\boldsymbol{\beta}}_n) = \hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)$.

Example 2 (S-scales). For $\mathbf{v} \in \mathbb{R}^p$, let $g_S(\mathbf{v})$ be the solution in s to the equation

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{r_i(\mathbf{v})}{s}\right) = b. \tag{5}$$

The S-scale estimator is now defined as

$$\hat{\sigma}_n^S = \inf_{\mathbf{v}} g_S(\mathbf{v}),$$

see Rousseeuw and Yohai (1984). The corresponding minimizer is the well-known regression S-estimator, that is, $\hat{\boldsymbol{\beta}}_n^S = \operatorname{argmin}_{\mathbf{v}} g_S(\mathbf{v})$. Note that the S-scale estimator $\hat{\sigma}_n^S$ is an M-scale estimator for the particular choice $\hat{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_n^S$, that is $\hat{\sigma}_n^M(\hat{\boldsymbol{\beta}}_n^S) = \hat{\sigma}_n^S$. It is well-known that the loss function ρ in (4) and (5) can be chosen so that $\hat{\sigma}_n^M(\hat{\boldsymbol{\beta}}_n)$ and $\hat{\sigma}_n^S$ are either robust or efficient. That is, there is a trade-off between robustness and efficiency for M and S-scale estimators.

Example 3 (One-step M-scales). To avoid a robustness-efficiency trade-off one can use the following procedure based on two loss functions ρ_0 and ρ. Given a preliminary robust regression estimator $\tilde{\boldsymbol{\beta}}_n$ we construct a highly robust M-scale estimator $\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)$ using a loss function ρ_0 in (4) that is tuned for robustness. Then, this robust scale estimator is used as initial scale in (2). That is, we define the one-step M-scale by

$$\hat{\sigma}_n^{M1}(\tilde{\boldsymbol{\beta}}_n) = \hat{\sigma}_n(\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n), \tilde{\boldsymbol{\beta}}_n) = \hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)\sqrt{\frac{1}{b\,n}\sum_{i=1}^{n}\rho\left(\frac{r_i(\tilde{\boldsymbol{\beta}}_n)}{\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)}\right)}.$$

It turns out that $\hat{\sigma}_n^{M1}(\tilde{\boldsymbol{\beta}}_n)$ inherits the robustness (BDP) of the initial scale estimator $\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)$ (determined by ρ_0). Hence, the loss function ρ can be chosen to obtain any desired efficiency without affecting the BDP of $\hat{\sigma}_n^{M1}(\tilde{\boldsymbol{\beta}}_n)$.

Example 4 ($\boldsymbol{\tau}$-scales). For $\mathbf{v} \in \mathbb{R}^p$, let

$$g_\tau(\mathbf{v}) = g_S(\mathbf{v}) \sqrt{\frac{1}{b\,n} \sum_{i=1}^n \rho\left(\frac{r_i(\mathbf{v})}{g_S(\mathbf{v})}\right)},$$

where $g_S(\mathbf{v})$ is defined by (5) with $\rho = \rho_0$. Then, the τ-scale estimator is defined as

$$\hat{\sigma}_n^\tau = \inf_{\mathbf{v}} g_\tau(\mathbf{v}),$$

see Yohai and Zamar (1988). The corresponding minimizer is the well-known regression τ-estimator. That is, $\hat{\boldsymbol{\beta}}_n^\tau = \operatorname{argmin}_{\mathbf{v}} g_\tau(\mathbf{v})$. Note that the τ-scale $\hat{\sigma}_n^\tau$ is a one-step M-scale for the particular choice $\hat{\boldsymbol{\beta}}_n = \hat{\boldsymbol{\beta}}_n^\tau$.

Example 5 (Extended one-step M-scales). The one-step M-scales in Example 3 updates the initial scale estimator (which is based on a preliminary robust regression estimator). For example, the initial scale estimator in Example 3 could be a highly robust S-estimator based on its companion highly robust but inefficient regression S-estimator. One may also consider incorporating a more efficient regression estimator. Let $\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)$ be a robust M-scale estimator corresponding to a preliminary regression estimator $\tilde{\boldsymbol{\beta}}_n$ and let $\hat{\boldsymbol{\beta}}_n$ be a more efficient regression estimator. We then define the extended one-step M-scale estimator as follows:

$$\hat{\sigma}^{EM1}(\tilde{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\beta}}_n) = \hat{\sigma}_n(\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n), \hat{\boldsymbol{\beta}}_n) = \hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n) \sqrt{\frac{1}{b\,n} \sum_{i=1}^n \rho\left(\frac{r_i(\hat{\boldsymbol{\beta}}_n)}{\hat{\sigma}_n^M(\tilde{\boldsymbol{\beta}}_n)}\right)}. \tag{6}$$

Example 6 (MM-scales). For $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$, let

$$g_{MM}(\mathbf{v}_1, \mathbf{v}_2) = g_S(\mathbf{v}_1) \sqrt{\frac{1}{b\,n} \sum_{i=1}^n \rho\left(\frac{r_i(\mathbf{v}_2)}{g_S(\mathbf{v}_1)}\right)},$$

where $g_S(\mathbf{v}_1)$ is again given by (5) with $\rho = \rho_0$. Then, we define the MM-scale as

$$\hat{\sigma}_n^{MM} = \inf_{\mathbf{v}_1, \mathbf{v}_2} g_{MM}(\mathbf{v}_1, \mathbf{v}_2).$$

Note that this estimator is a special case of (6) and thus of (2). Indeed, let $\hat{\boldsymbol{\beta}}_n^S$ be the minimizer of $g_S(\mathbf{v}_1)$ as in Example 2 and let $\hat{\boldsymbol{\beta}}_n^{MM}$ be the regression MM-estimator of Yohai (1987) which is the minimizer in $\mathbf{v}_2$ of

$$\hat{\sigma}_n^M(\hat{\boldsymbol{\beta}}_n^S) \sqrt{\frac{1}{b\,n} \sum_{i=1}^n \rho\left(\frac{r_i(\mathbf{v}_2)}{\hat{\sigma}_n^M(\hat{\boldsymbol{\beta}}_n^S)}\right)}.$$

If the loss function satisfies $2\rho(u) - \rho'(u)u \geq 0$, then

$$\hat{\sigma}_n^{MM} = \hat{\sigma}_n^M(\hat{\boldsymbol{\beta}}_n^S) \sqrt{\frac{1}{b\,n} \sum_{i=1}^n \rho\left(\frac{r_i(\hat{\boldsymbol{\beta}}_n^{MM})}{\hat{\sigma}_n^M(\hat{\boldsymbol{\beta}}_n^S)}\right)}.$$

There is an important difference between Examples 1, 3, and 5 and Examples 2, 4 and 6. In Examples 1, 3 and 5, the scale estimator is based on one or two *preliminary* regression estimators. Such preliminary regression estimators are not required in Examples 2, 4 and 6, where the scale and regression estimators are obtained *simultaneously*.

The rest of the paper is organized as follows. In Section 2 we derive the influence function, asymptotic variance and the gross-error sensitivity of the generalized M-scales defined in (2). In Section 3 we investigate their maxbias behavior. In particular, we determine minimax-bias choices for $\hat{\boldsymbol{\beta}}_n$ and $\tilde{\boldsymbol{\beta}}_n$. In Section 4 we solve a Hampel-like optimality problem. We find combinations of loss functions such that the scale estimator has maximal efficiency subject to a constraint on the breakdown point and gross-error sensitivity. In Section 5 we make a comparison between τ and MM-estimators taking the maxbias of both regression and scale estimators into account. Section 6 is an Appendix that collects some proofs and technical details.

2. Influence function and efficiency

The influence function (IF) of a functional T at the central distribution H_0 measures the effect on the functional of an infinitesimal amount of contamination placed at a single point $(y_0, \mathbf{z}_0)$ (see e.g. Hampel et al., 1986). Let us denote $H_\epsilon^{(y_0,\mathbf{z}_0)} = (1-\epsilon)H_0 + \epsilon\Delta_{(y_0,\mathbf{z}_0)}$, where $\Delta_{(y_0,\mathbf{z}_0)}$ denotes the point-mass distribution that puts all of its mass at $(y_0, \mathbf{z}_0)$. Then, the influence function of the functional T at $(y_0, \mathbf{z}_0)$ and central model H_0 is defined as

$$IF(y_0, \mathbf{z}_0; T, H_0) = \lim_{\epsilon\to 0} \frac{T(H_\epsilon^{(y_0,\mathbf{z}_0)}) - T(H_0)}{\epsilon} = \left.\frac{\partial T(H_\epsilon^{(y_0,\mathbf{z}_0)})}{\partial \epsilon}\right|_{\epsilon=0}.$$

The notation $\left.\frac{\partial T(H_\epsilon^{(y_0,\mathbf{z}_0)})}{\partial \epsilon}\right|_{\epsilon=0}$ means that the function $T(H_\epsilon^{(y_0,\mathbf{z}_0)})$ is first differentiated and then the derivative is evaluated at $\epsilon = 0$. We now consider generalized M-scale functionals $\hat{\sigma}(H) = \hat{\sigma}(H; \hat{s}, \hat{\boldsymbol{\beta}})$ given by (3), where $\hat{\boldsymbol{\beta}}(H)$ is a regression functional corresponding to $\hat{\boldsymbol{\beta}}_n$ and $\hat{s}(H)$ is a scale functional corresponding to the initial scale estimator $\hat{s}_n$. We assume that $\hat{\boldsymbol{\beta}}(H)$ is an affine, regression and scale equivariant regression functional (see e.g. Rousseeuw and Leroy, 1987, p. 116) which is Fisher-consistent at H_0, i.e. $\hat{\boldsymbol{\beta}}(H_0) = \boldsymbol{\beta}_0$. Moreover, the initial scale functional $\hat{s}(H)$ is assumed to be scale equivariant and Fisher-consistent at H_0, i.e. $\hat{s}(H_0) = \sigma_0$. For the choice $b = E_{F_0}\{\rho(u)\}$, it then immediately follows from (3) that the resulting generalized M-scale functional $\hat{\sigma}(H; \hat{s}, \hat{\boldsymbol{\beta}})$ is Fisher-consistent at H_0, i.e. $\hat{\sigma}(H_0; \hat{s}, \hat{\boldsymbol{\beta}}) = \sigma_0$. Note that by equivariance we can assume without loss of generality that $\boldsymbol{\beta}_0 = \mathbf{0}$ and $\sigma_0 = 1$.

We have the following result.

Theorem 1. *Suppose that assumption (A1) holds for the loss function ρ. Moreover, suppose that the functionals $\hat{\boldsymbol{\beta}}(H)$ and $\hat{s}(H)$ are Fisher consistent at*

$H_0 = (G_0, F_0)$. *Then, the influence function of the generalized M-scale functional* $\hat{\sigma}(H) = \hat{\sigma}(H; \hat{s}, \hat{\boldsymbol{\beta}})$ *and the influence function of the initial scale functional* $\hat{s}(H)$ *satisfy the following relation*

$$IF(y_0, \mathbf{z}_0; \hat{\sigma}, H_0) = \left(1 - \frac{E_{F_0}[\psi(u)\, u]}{2b}\right) IF(y_0, \mathbf{z}_0; \hat{s}, H_0) + \frac{\rho(y_0) - b}{2b}. \quad (7)$$

From (7) we notice that the influence function of $\hat{\sigma}(H; \hat{s}, \hat{\boldsymbol{\beta}})$ does not depend on the Fisher-consistent regression functional $\hat{\boldsymbol{\beta}}(H)$. On the other hand, it does critically depend on the choice of the initial scale functional $\hat{s}(H)$ through its influence function.

We know from Example 1 that in the case of M-scale estimators the corresponding functional can be written as $\hat{\sigma}^M(H; \tilde{\boldsymbol{\beta}}) = \hat{\sigma}(H; \hat{s}, \tilde{\boldsymbol{\beta}})$ with $\hat{s}(H) = \hat{\sigma}^M(H; \tilde{\boldsymbol{\beta}})$. Therefore, in this case $IF(y_0, \mathbf{z}_0; \hat{\sigma}, H_0) = IF(y_0, \mathbf{z}_0; \hat{s}, H_0)$ and both are equal to $IF(y_0, \mathbf{z}_0; \hat{\sigma}^M, H_0)$. Using this in (7) and solving for $IF(y_0, \mathbf{z}_0; \hat{\sigma}^M, H_0)$ we obtain the well known result

$$IF(y_0, \mathbf{z}_0; \hat{\sigma}^M, H_0) = IF(y_0; \hat{\sigma}^M, F_0) = \frac{\rho(y_0) - b}{E_{F_0}[\psi(u)\, u]}. \quad (8)$$

In this case the influence function of $\hat{\sigma}^M(H; \tilde{\boldsymbol{\beta}})$ also does not depend on the distribution G_0 of the regressors (see also Van Aelst and Willems, 2005).

It can be shown that under regularity conditions, generalized M-scales are asymptotically normal with asymptotic variance obtained by integrating the square of the IF (see e.g. Hampel et al., 1986). That is

$$\mathrm{ASV}(\hat{\sigma}, H_0) = E_{H_0}[IF^2(y, \mathbf{z}; \hat{\sigma}, H_0)].$$

If the scale estimator $(\hat{\sigma}_n(\hat{s}_n, \hat{\boldsymbol{\beta}}_n)$ is only based on residuals w.r.t. one or two consistent regression estimators, then the asymptotic distribution of the scale estimator at the model depends only on the error distribution F_0. Hence, in this case the efficiency of the scale estimator becomes independent of the design and the above expression for the asymptotic variance remains valid. Moreover, the gross-error sensitivity of the generalized M-scale functional $\hat{\sigma}(H; \hat{s}, \hat{\boldsymbol{\beta}})$ is given by

$$\mathrm{GES}(\hat{\sigma}, H_0) = \sup_{y, \mathbf{z}} |IF(y, \mathbf{z}; \hat{\sigma}, H_0)|.$$

Extended one-step M-scales. Now, we consider the important particular case of Example 5 where the initial scale estimator $\hat{s}(H)$ is an M-scale estimator using a highly robust loss function ρ_0 (with corresponding constant b_0 to obtain consistency at F_0) that satisfies assumption (A1) and based on residuals formed by using a highly robust, Fisher-consistent preliminary regression estimator $\tilde{\boldsymbol{\beta}}_n$. Hence, the corresponding functional depends on two regression estimators and is denoted by $\hat{\sigma}^{EM1}(H; \tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}})$. Combining (7) and (8) we obtain that

$$\begin{aligned} IF(y_0, \mathbf{z}_0; \hat{\sigma}^{EM1}, H_0) &= IF(y_0; \hat{\sigma}^{EM1}, F_0) \\ &= \frac{1}{2b}\left[W_{F_0}(\rho_0(y_0) - b_0) + (\rho(y_0) - b)\right] \end{aligned} \quad (9)$$

with

$$W_{F_0} = \frac{2b - E_{F_0}\{\psi(u)u\}}{E_{F_0}\{\psi_0(u)u\}}. \tag{10}$$

Note that (9) reduces to (8) when $\rho_0 = \rho$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$ because $\hat{\sigma}^{EM1}(H_0; \tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}) = \hat{\sigma}^M(H_0; \hat{\boldsymbol{\beta}})$ in this case. Moreover, as before, the influence function of $\hat{\sigma}^{EM1}(H; \tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}})$ does not depend on the Fisher-consistent regression functionals $\hat{\boldsymbol{\beta}}(H)$ and $\tilde{\boldsymbol{\beta}}(H)$ nor on the distribution function G_0 of the explanatory variables. Hence, not surprisingly, it follows immediately from (7) that if a common M-scale is used as initial scale, then for a fixed loss function ρ the resulting generalized M-scale functionals all have the same influence function. Note that one-step M-scale, τ-scale, and MM-scale functionals all are special cases of extended one-step M-scale functionals, corresponding to particular choices of the auxiliary regression functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$. Therefore, for a fixed pair of loss functions ρ_0 and ρ, the corresponding one-step M-scale, τ-scale, extended one-step M-scale, and MM-scale functionals all have the same influence function, given by (9). Hence, they also have the same gross-error sensitivity and asymptotic efficiency.

Yohai and Zamar (1988) showed that the τ-estimator for the regression coefficients asymptotically behaves as a regression M-estimator with loss function ρ_τ which is a weighted average between ρ and ρ_0. Here, we show a similar result for the one-step M-scales and the extended one-step M-scales. Consider the loss function ρ_τ given by

$$\rho_\tau(t) := W_{F_0}\,\rho_0(t) + \rho(t) \qquad \text{for } t \in \mathbb{R}, \tag{11}$$

with W_{F_0} defined in (10). Now, we can define the scale M-estimator $\hat{\sigma}_n^{M,\tau}$ as the solution in s to

$$\frac{1}{n}\sum_{i=1}^{n} \rho_\tau\left(\frac{r_i(\boldsymbol{\beta}_0)}{s}\right) = b_\tau, \tag{12}$$

with $b_\tau = E_{F_0}\{\rho_\tau(u)\}$. The influence function of the M-scale functional $\hat{\sigma}^{M,\tau}(H)$ can be obtained immediately by setting $\rho = \rho_\tau$ in (8). It can easily be seen that this influence function is equal to the right hand side of (9). Hence, we have the following corollary.

Corollary 1. *Suppose that the loss functions ρ and ρ_0 satisfy assumption (A1). Moreover, suppose that the regression functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$ are Fisher consistent. Then, the M-scale functional $\hat{\sigma}^{M,\tau}(H)$ defined by (12) has the same influence function, gross-error sensitivity and asymptotic variance as the one-step M-scale functional $\hat{\sigma}^{M1}(H; \hat{\boldsymbol{\beta}})$ and the extended one-step M-scale functional $\hat{\sigma}^{EM1}(H; \tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}})$.*

Note that there is an important difference between the M-scale $\hat{\sigma}^{M,\tau}(H)$ on the one hand and the [extended] one-step M-scales [$\hat{\sigma}^{EM1}(H; \tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}})$] $\hat{\sigma}^{M1}(H; \hat{\boldsymbol{\beta}})$ on the other hand. Namely, while the breakdown point of the [extended] one-step M-scales is given by $\min\{b_0, 1-b_0, bdp(\hat{\boldsymbol{\beta}}), bdp(\tilde{\boldsymbol{\beta}})\}$, the breakdown point of the M-scale $\hat{\sigma}^{M,\tau}(H)$ is given by $\min\{b_\tau, 1-b_\tau\}$. Therefore, the [extended] one-step M-scale can be tuned to be simultaneously robust and efficient while the M-scale can be tuned to achieve only one of these two features at the time. Note that it follows from Corollary 1 that [extended] one-step M-scales have a bounded influence function. The S, τ and MM-scales thus have a bounded influence function, unlike their associated high-breakdown regression functionals whose influence function is unbounded for good leverage points (i.e. points whose $\mathbf{z}_i$ is outlying, but still follow the regression model).

3. Maxbias and optimal choice of the regression estimators

Consider a fixed pair of loss functions ρ_0 and ρ, chosen to obtain a desired BDP and efficiency for the corresponding generalized M-scale estimators (2). In the previous section, we showed that the influence function of these generalized M-scale estimators does not depend on the choice of the regression estimators $\tilde{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\beta}}_n$. Therefore, the choice of the regression estimators $\tilde{\boldsymbol{\beta}}_n$ and $\hat{\boldsymbol{\beta}}_n$ in (2) can not be guided by the comparison of the BDP, GES or asymptotic variance of the resulting scale estimators. In view of that, a reasonable approach for choosing the robust regression estimators is to examine the actual maxbias behavior of the corresponding scale estimators. Throughout this section we assume without loss of generality that $\lim_{t\to\infty}\rho(t) = \lim_{t\to\infty}\rho_0(t) = 1$.

It is well known that the bias of a scale functional can be of two distinct types: (a) explosion bias, which is due to overestimation caused by *outliers*, and (b) implosion bias, which is due to underestimation caused by *inliers*. We denote by $\hat{\sigma}^+(\epsilon;\hat{s},\hat{\boldsymbol{\beta}})$ and $\hat{\sigma}^-(\epsilon;\hat{s},\hat{\boldsymbol{\beta}})$ the supremum and infimum over $\mathcal{H}_\epsilon$ of the generalized M-scale functional $\hat{\sigma}(H;\hat{s},\hat{\boldsymbol{\beta}})$. Following Martin et al. (1989), the overall (asymptotic) bias can then be measured by a generalized bias function which penalizes explosion and implosion bias on different scales. For example, when penalizing both biases on a logarithmic scale the maximum generalized bias becomes

$$\begin{aligned}\overline{B}(\epsilon;\hat{s},\hat{\boldsymbol{\beta}}) &= \sup_{H_\epsilon\in\mathcal{H}_\epsilon}\left[\max\left\{\log\left(\frac{\sigma(H_\epsilon;\hat{s},\hat{\boldsymbol{\beta}})}{\sigma_0}\right), -\log\left(\frac{\sigma(H_\epsilon;\hat{s},\hat{\boldsymbol{\beta}})}{\sigma_0}\right)\right\}\right] \\ &= \max\ \left\{\log\left(\frac{\sigma^+(\epsilon;\hat{s},\hat{\boldsymbol{\beta}})}{\sigma_0}\right), -\log\left(\frac{\sigma^-(\epsilon;\hat{s},\hat{\boldsymbol{\beta}})}{\sigma_0}\right)\right\}. \end{aligned} \tag{13}$$

We now consider the maxbias behavior of generalized M-scale functionals $\sigma(H;\hat{s},\hat{\boldsymbol{\beta}})$ with initial scale functional $\hat{s}\,(H) = \hat{s}(H;\tilde{\boldsymbol{\beta}})$ based on residuals $r(\tilde{\boldsymbol{\beta}}(H))$. In particular, we investigate which choice of the regression functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$ minimizes the maxbias of $\sigma(H;\hat{s}(H;\tilde{\boldsymbol{\beta}}),\hat{\boldsymbol{\beta}})$. Note that we will sometimes use the notations $\hat{s}(\tilde{\boldsymbol{\beta}})$ or $\hat{s}(H;\tilde{\boldsymbol{\beta}})$ to emphasize the dependence of $\hat{s}\,(H)$ on the regression residuals $r(\tilde{\boldsymbol{\beta}}(H))$.

Due to regression, affine and scale equivariance of our functionals we can assume without loss of generality that $\boldsymbol{\beta}_0 = \mathbf{0}$ and $\sigma_0 = 1$. Furthermore, we require that the initial scale functional $\hat{s}(H;\tilde{\boldsymbol{\beta}})$ satisfies the following assumption

(A4) For any $\mathbf{v} \in \mathbb{R}^p$ and $H_\epsilon \in \mathcal{H}_\epsilon$ we have that

(i) $\hat{s}(H_\epsilon; \mathbf{v}) \geq \hat{s}(H_\epsilon^0; \mathbf{0})$,

(ii) $\hat{s}(H_\epsilon; \mathbf{0}) \leq \hat{s}(H_\epsilon^\infty; \mathbf{v})$,

where $H_\epsilon^0 = (1-\epsilon)H_0 + \epsilon\Delta_{(0,\mathbf{0})}$ and $H_\epsilon^\infty = (1-\epsilon)H_0 + \epsilon\Delta_{(\infty,\mathbf{0})}$.

Assumption (A4) is a technical condition and specifies that the robust scale estimator should asymptotically behave in the "expected way" in the presence of point mass contamination at zero and infinity. Scale M, S and least trimmed of squares (LTS) estimators are some typical choices for the initial scale estimator in (2). M and S-estimators are described in Examples 1 and 2 above and scale LTS estimators are described in the Appendix. Lemmas 1 and 2 in the Appendix show that scale LTS, M-, and S- estimators satisfy condition (A4).

Theorem 2. *Suppose that assumptions (A1)-(A3) hold. Moreover, suppose that $\hat{s}(H; \mathbf{v})$ is a scale functional that satisfies condition (A4). For any $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$, consider the function*

$$g_{GM}(H; \mathbf{v}_1, \mathbf{v}_2) = \hat{s}(H; \mathbf{v}_1)\sqrt{\frac{1}{b}E_H\left\{\rho\left(\frac{r_i(\mathbf{v}_2)}{\hat{s}(H; \mathbf{v}_1)}\right)\right\}},$$

and define the regression functionals $\tilde{\boldsymbol{\beta}}^{opt}(H)$ and $\hat{\boldsymbol{\beta}}^{opt}(H)$ by

$$(\tilde{\boldsymbol{\beta}}^{opt}(H), \hat{\boldsymbol{\beta}}^{opt}(H)) = \underset{\mathbf{v}_1, \mathbf{v}_2}{argmin}\, g_{GM}(H; \mathbf{v}_1, \mathbf{v}_2).$$

Then, for any regression, scale and affine equivariant functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$ which satisfy $\tilde{\boldsymbol{\beta}}(H_\epsilon^0) = \hat{\boldsymbol{\beta}}(H_\epsilon^0) = \mathbf{0}$, it holds that

$$\sigma^+(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}^{opt}), \hat{\boldsymbol{\beta}}^{opt}) \leq \sigma^+(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}})$$

$$\sigma^-(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}^{opt}), \hat{\boldsymbol{\beta}}^{opt}) = \sigma^-(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}).$$

Note that for any regression functional $\hat{\boldsymbol{\beta}}(H)$ the condition $\hat{\boldsymbol{\beta}}(H_\epsilon^0) = \mathbf{0}$ simply means that the regression functional is not affected by contamination that consists of inliers lying exactly on the true regression hyperplane (determined by $\boldsymbol{\beta}_0 = \mathbf{0}$). This is natural for regression functionals that are based on residuals.

It immediately follows from Theorem 2 that $\overline{B}(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}^{\text{opt}}), \hat{\boldsymbol{\beta}}^{\text{opt}}) \leq \overline{B}(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}})$ (see (13)) for regression functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$. Hence, Theorem 2 shows that the maxbias of the scale estimator is minimized when the regression and scale parameters are estimated simultaneously. It can be shown that Theorem 2 still holds if contamination can only occur in the error distribution F_0. Hence, the result that the maxbias of the scale estimator is minimal when the regression parameters are estimated simultaneously is rather general and independent of the actual design for the regression model.

Corollary 2. *From the proof of Theorem 2 it follows that the minimax bias estimators reach their maximal explosion bias at $H = H_\epsilon^\infty$ and their minimal implosion bias at $H = H_\epsilon^0$. Moreover, in both cases the corresponding regression estimators take the value $\mathbf{0}$. Using (3) we can thus obtain explicit expressions for the maxbias of the optimal generalized M-scale in terms of the explosion maxbias $s^+(\epsilon; \tilde{\boldsymbol{\beta}}^{opt}) = s(H_\epsilon^\infty; \mathbf{0})$ and implosion maxbias $s^-(\epsilon; \tilde{\boldsymbol{\beta}}^{opt}) = s(H_\epsilon^0; \mathbf{0})$ of the initial residual scale estimator. That is,*

$$\sigma^+(\epsilon; s(\tilde{\boldsymbol{\beta}}^{opt}), \hat{\boldsymbol{\beta}}^{opt}) = \frac{s^+(\epsilon; \tilde{\boldsymbol{\beta}}^{opt})}{\sqrt{b}} \sqrt{(1-\epsilon) E_{F_0}\left\{\rho\left(\frac{u}{s^+(\epsilon; \tilde{\boldsymbol{\beta}}^{opt})}\right)\right\} + \epsilon} \tag{14}$$

$$\sigma^-(\epsilon; s(\tilde{\boldsymbol{\beta}}^{opt}), \hat{\boldsymbol{\beta}}^{opt}) = \frac{s^-(\epsilon; \tilde{\boldsymbol{\beta}}^{opt})}{\sqrt{b}} \sqrt{(1-\epsilon) E_{F_0}\left\{\rho\left(\frac{u}{s^-(\epsilon; \tilde{\boldsymbol{\beta}}^{opt})}\right)\right\}} \tag{15}$$

Lemma 2 in the Appendix shows that the M-scale functional $\hat{\sigma}^M(H; \tilde{\boldsymbol{\beta}})$ corresponding to (4) satisfies condition (A4). Therefore, we have the following corollary.

Corollary 3 (MM-scale optimality)**.** *Suppose that assumptions (A1)-(A3) hold. Then, for any regression, scale and affine equivariant functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$ which satisfy $\tilde{\boldsymbol{\beta}}(H_\epsilon^0) = \hat{\boldsymbol{\beta}}(H_\epsilon^0) = \mathbf{0}$, it holds that*

$$\sigma_{MM}^+(\epsilon) \le \sigma_{EM1}^+(\epsilon; \tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}})$$

$$\sigma_{MM}^-(\epsilon) = \sigma_{EM1}^-(\epsilon; \tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}).$$

where $\sigma_{MM}^+(\epsilon)$ and $\sigma_{MM}^-(\epsilon)$ are the explosion and implosion maxbiases of the corresponding scale MM-estimator.

Corollary 3 shows that the optimal regression estimators, yielding the minimax bias scale estimator, are the regression S and MM-estimators respectively. That is,

$$\sigma_{MM}^+(\epsilon) = \sigma_{EM1}^+\left(\epsilon; \hat{\boldsymbol{\beta}}^S, \hat{\boldsymbol{\beta}}^{MM}\right) \tag{16}$$

$$\sigma_{MM}^-(\epsilon) = \sigma_{EM1}^-\left(\epsilon; \hat{\boldsymbol{\beta}}^S, \hat{\boldsymbol{\beta}}^{MM}\right). \tag{17}$$

Remark 1. (τ and S-scale Optimality) *By imposing the restriction $\tilde{\boldsymbol{\beta}}(H) = \hat{\boldsymbol{\beta}}(H)$ it directly follows from the proof of Corollary 3 that*

$$\sigma_\tau^+(\epsilon) = \sigma_{M1}^+(\epsilon; \hat{\boldsymbol{\beta}}^\tau) \le \sigma_{M1}^+(\epsilon; \hat{\boldsymbol{\beta}})$$
$$\sigma_\tau^-(\epsilon) = \sigma_{M1}^-(\epsilon; \hat{\boldsymbol{\beta}}^\tau) = \sigma_{M1}^-(\epsilon; \hat{\boldsymbol{\beta}})$$

for any regression, scale and affine equivariant functional $\hat{\boldsymbol{\beta}}(H)$ satisfying the conditions of Theorem 2.

Since M-scales are a particular case of one-step M-scales (with $\rho_0 = \rho$), this result also shows that the S-scale has minimax bias within the class of M-scales. That is,

$$\sigma_S^+(\epsilon) = \sigma_M^+(\epsilon; \hat{\boldsymbol{\beta}}^S) \leq \sigma_M^+(\epsilon; \hat{\boldsymbol{\beta}}),$$

$$\sigma_S^-(\epsilon) = \sigma_M^-(\epsilon; \hat{\boldsymbol{\beta}}^S) = \sigma_M^-(\epsilon; \hat{\boldsymbol{\beta}}),$$

for any regression, scale and affine equivariant functional $\hat{\boldsymbol{\beta}}(H)$ satisfying the conditions of Theorem 2.

Corollary 3 and Remark 1 show that the S-scale $\hat{\sigma}^S$ of Rousseeuw and Yohai (1984), the τ-scale $\hat{\sigma}^\tau$ of Yohai and Zamar (1988) and the MM-scale $\hat{\sigma}^{MM}$ are optimal in the sense that they minimize the maximal possible bias of the scale functional within their respective classes.

From Corollary 2 together with (4) we can see that the explosion and implosion maxbiases of an S-scale estimator are implicitly determined by the equations

$$E_{F_0}\left\{\rho_0\left(\frac{u}{\sigma_S^+(\epsilon)}\right)\right\} = \frac{b_0 - \epsilon}{1-\epsilon}, \tag{18}$$

and

$$E_{F_0}\left\{\rho_0\left(\frac{u}{\sigma_S^-(\epsilon)}\right)\right\} = \frac{b_0}{1-\epsilon}, \tag{19}$$

respectively

By combining (14)-(15) with (16)-(17) and Remark 1, it follows that if the same loss functions ρ_0 and ρ are used for the τ and MM-scale functionals, then

$$\sigma_{MM}^+(\epsilon) = \sigma_\tau^+(\epsilon) = \frac{\sigma_S^+(\epsilon)}{\sqrt{b}}\sqrt{(1-\epsilon)E_{F_0}\left\{\rho\left(\frac{u}{\sigma_S^+(\epsilon)}\right)\right\} + \epsilon} \tag{20}$$

$$\sigma_{MM}^-(\epsilon) = \sigma_\tau^-(\epsilon) = \frac{\sigma_S^-(\epsilon)}{\sqrt{b}}\sqrt{(1-\epsilon)E_{F_0}\left\{\rho\left(\frac{u}{\sigma_S^-(\epsilon)}\right)\right\}} \tag{21}$$

with $\sigma_S^+(\epsilon)$ and $\sigma_S^-(\epsilon)$ given by (18)-(19) respectively. Note that from (18)-(21) it follows immediately that the maxbiases of S, τ and MM-scales only depend on the error distribution F_0 but are independent of the distribution G_0 of the predictors.

Table 1 gives an overview of classes of loss functions $\rho(t)$ that have been used for high breakdown M-scales. Two loss functions do not fit in this table. These are the Yohai-Zamar optimal loss function for regression τ-estimators (Yohai and Zamar, 1997):

$$\rho_c(t) = \begin{cases} 1.38t^2/c^2 & |t/c| \leq 2/3 \\ .55 - 2.69(\frac{t}{c})^2 + 10.76(\frac{t}{c})^4 - 11.66(\frac{t}{c})^6 + 4.04(\frac{t}{c})^8 & 2/3 < |t/c| \leq 1 \\ 1 & |t/c| > 1, \end{cases} \tag{22}$$

Table 1: Robustness-efficiency trade-off for M-scales: Efficiency for 50% BDP M-scale functionals for several choices of the loss function ρ.

Name	$\rho_c(t)$	c	Eff (%)
Step function	$I(\|t\| \geq c)$	0.674	36.7
Huber	$\min(t^2/c^2, 1)$	1.041	50.6
Yohai-Zamar	see (22)	1.212	50.9
Biweight	$\min(3(\frac{t}{c})^2 - 4(\frac{t}{c})^4 + (\frac{t}{c})^6, 1)$	1.547	53.9
Welsh	$1 - \exp(-t^2/c^2)$	0.666	54.9
Cauchy	$t^2/(t^2 + c^2)$	0.612	52.2
Truncated linear	$\min(\|t/c\|, 1)$	1.470	61.5
Croux	see (23)	0.01	92.9

and the Croux loss function (Croux, 1994):

$$\rho_c(t) = \begin{cases} (c + b - k)\sqrt{y^2/c} & 0 \leq y^2 < c \\ 1 - (k - y^2)/a & c \leq y^2 < k \\ 1 & y^2 \geq k \end{cases} \tag{23}$$

We now examine the (generalized) maxbias of S-scale functionals as defined in (13) when the error distribution is the standard Gaussian distribution. Remember that the maxbias of S-scales depends only on F_0, so G_0 doesn't need to be specified. From (18) and (19) it can immediately be seen that the maxbias of S-scales goes to infinity as $\epsilon \to \min\{b_0, 1 - b_0\}$. Hence b_0 is set to 0.5 to obtain scale functionals with maximal breakdown point. For each class of loss functions, the constant c must then be chosen to guarantee that $b_0 = 0.5 = E_\Phi\{\rho_c(u)\}$. However, for all the commonly used loss functions the constant c also determines the efficiency of the S-scale functional. This leads to a trade-off between robustness and efficiency for S-scale functionals, resulting in a low efficiency as can be seen from the last column of Table 1. On the other hand, if the constant c is tuned to obtain high efficiency, then this results in a low BDP. An exception to this efficiency-BDP trade-off is the Croux loss function which can be tuned to simultaneously obtain maximal BDP and high efficiency. However, as shown below in Figure 1, there is a severe efficiency/maxbias tradeoff for this scale estimator.

It follows from (14)-(15) that the maxbias of minimax bias generalized M-scale functionals remains bounded whenever the initial scale functional does not break down. Hence, minimax bias generalized M-scales inherit the BDP of the initial scale estimator. Therefore, the constant c can be determined to obtain a highly efficient scale estimator without affecting its BDP. However, this does not imply that the increased efficiency can not have a high effect on the maximal possible bias of the scale estimator. Therefore, we now investigate the effect of the increased efficiency on the maxbias of the scale functionals. In Figure 1 we compare the maxbiases of S-scales and τ/MM-scales. The top panel shows the overall maximum generalized bias of the scale functionals while the other two panels show the explosion and implosion maxbias curves separately, using the

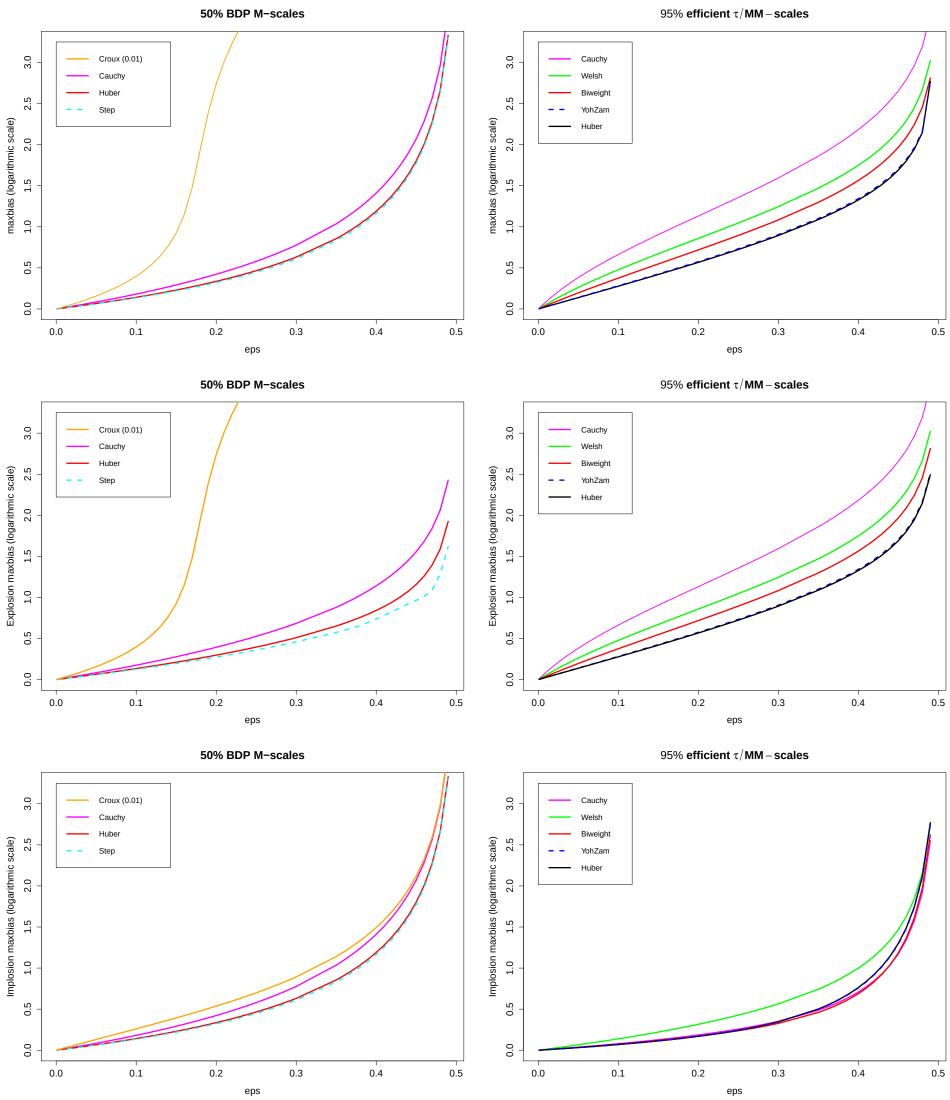

Figure 1: The left plots show maxbias curves for 50% BDP S-scale functionals. The right plots shows maxbias curves for 95% efficient τ/MM-scale functionals. The top panel shows the overall maximum generalized bias curves. The middle panel shows the explosion maxbias curves and the bottom panel shows the implosion maxbias curves. The legend of each plot describes the curves from top left to bottom right.

logarithmic scale in (13). The left plots in Figure 1 show the maxbias curves of the 50% breakdown point S-scales. The curves for the Yohai-Zamar, biweight and Welsh loss function are omitted because they are indistinguishable from the maxbias curve for the Huber truncated square loss function. The right panels in

Figure 1 show the maxbias curves for several 95% efficient τ/MM-scales. From the left panels we see that although the S-scale based on the Croux loss function can combine high efficiency with high BDP, this comes at a high price in terms of maxbias due to explosion bias. On the other hand, comparing the maxbias curves in the left and right panels reveals that the increased efficiency of the τ/MM-scales can come with only a modest effect on the maxbias compared to the maxbias of the initial S-scale, as can be seen for the Huber and Yohai-Zamar loss functions. For other error distributions F_0 such as Cauchy errors, similar results can be obtained.

The good maxbias behavior of τ/MM-scales is further illustrated in Figure 2 where we compare the maxbias curves for the 50% BDP S-scale (50.6% efficiency), the 95% efficient S-scale (17% BDP) and the τ/MM-scale (50% BDP, 95% efficiency) based on the Huber loss function. For the other loss functions, the plots are very similar. From this plot we can see that, as expected, increasing the efficiency has a large adverse effect on the (explosion) maxbias of the S-scale functional. On the other hand, for most contamination levels the maxbias of the τ/MM-scale is only slightly higher than the maxbias of the initial S-scale functional. For large fractions of contamination the maxbias of the τ/MM-scale is even lower than that of the initial S-scale functional, because τ/MM-scales are even more robust against inliers as can be seen in the bottom right plot of Figure 2.

4. Optimal loss functions

In this section we focus on extended one-step M-scale estimators. In Section 2 we derived the asymptotic variance and gross-error sensitivity of these estimators and showed that they do not depend on G_0 nor the regression estimators. However, they clearly depend on the choice of the loss functions ρ_0 and ρ. This motivates considering the following optimality problem. Given that the scale estimator is tuned to obtain a desired breakdown point $\bar{\epsilon}$, e.g. $\bar{\epsilon} = 50\%$, and given an upper bound $\bar{\gamma}$ on its gross-error sensitivity, we seek for optimal loss functions $\rho_0^\star$ and $\rho^\star$ that maximize the asymptotic Gaussian efficiency of the extended one-step M-scale estimator $\hat{\sigma}^{EM1}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}})$. That is, we wish to find a pair of optimal loss functions $(\rho_0^\star, \rho^\star)$ that solve the following problem:

$$(\rho_0^\star, \rho^\star) = \underset{\rho_0, \rho}{\operatorname{argmin}} \operatorname{ASV}(\hat{\sigma}^{EM1}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}), \Phi) \tag{24}$$

where the minimum is over all (ρ_0, ρ) such that

$$\operatorname{GES}(\hat{\sigma}^{EM1}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}), \Phi) \leq \bar{\gamma} \quad \text{and} \quad \bar{\epsilon}(\hat{\sigma}^{EM1}(\tilde{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}})) \geq \bar{\epsilon} \tag{25}$$

In Corollary 1 we showed that the extended one-step M-scales are asymptotically equivalent to the M-scale defined in (12) because they actually have the same influence function. Now, we consider loss functions $\rho_k(t)$ of the form

$$\rho_k(t) = \chi_k(t) - \chi_k(0), \text{ where } \chi_k(t) = [t^2 - 1 - a(k)]_{-k}^{k} \tag{26}$$

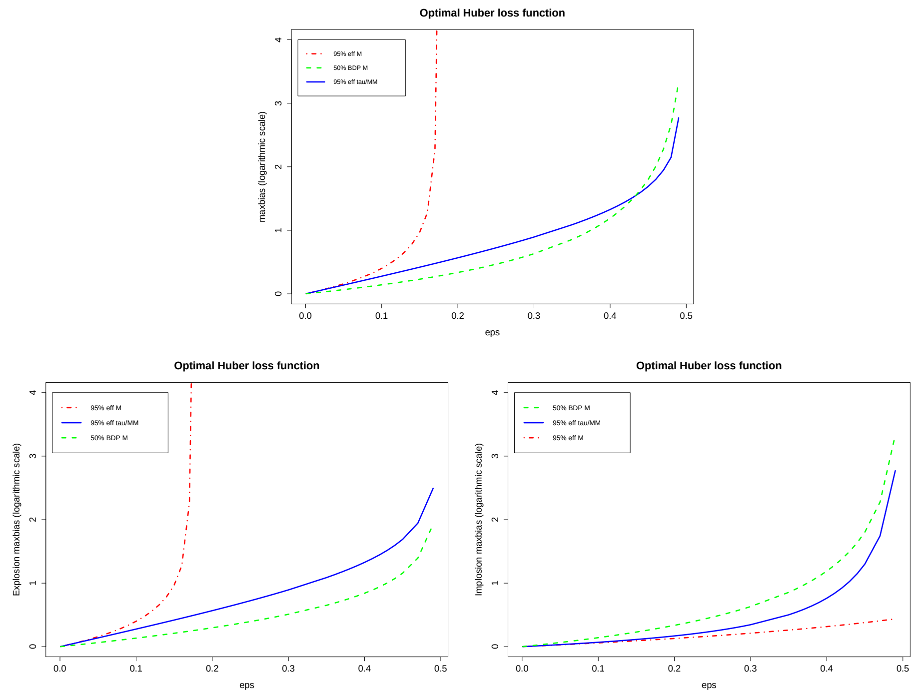

Figure 2: Maxbias curves for 50% BDP S-scale, 95% efficient S-scale and 95% efficient τ/MM-scale functionals based on Huber loss function. The top panel shows the overall maximum generalized bias curves. The plots in the bottom panel show the explosion (left) and implosion (right) maxbias separately.

with $a(k)$ chosen such that $E_\Phi[\chi_k(u)] = 0$. In (Hampel et al., 1986, p. 107 and 122) it is shown that the loss functions ρ_k given by (26) yield optimal M-scales in the sense that these estimators have maximal Gaussian efficiency subject to a bound on the GES. The bound on the GES can be any value between 1.167 (minimal possible GES, achieved by the MAD) and ∞. Hence, for any choice $\bar{\gamma} \in [1.167, \infty)$ there exists a corresponding $k(\bar{\gamma})$ such that the M-scale based on $\rho_{k(\bar{\gamma})}$ has GES $= \bar{\gamma}$ and minimizes the ASV among all M-scales with GES $\leq \bar{\gamma}$.

Given the formal asymptotic equivalence between extended one-step M-scales and M-scales, established in Corollary 1, it is clear that if we can find a pair of loss functions $(\rho_0^\star, \rho^\star)$ with $E_\Phi\{\rho_0^\star(u)\} = \bar{\epsilon}$ such that the corresponding $\rho_\tau^\star(t) = W_\Phi\, \rho_0^\star(t) + \rho^\star(t)$ as defined in (11) with $W_F(\Phi)$ given by (10) satisfies $\rho_\tau^\star = \rho_{k(\bar{\gamma})}$, then this pair $(\rho_0^\star, \rho^\star)$ is a solution to (24)-(25). The following theorem shows how to obtain such optimal pairs.

Theorem 3. *Given a minimal breakdown point $\bar{\epsilon}$ and a maximal gross-error-sensitivity $\bar{\gamma}$, let $k = k(\bar{\gamma})$ be such that ρ_k maximizes the M-scale efficiency subject to the given GES bound $\bar{\gamma}$. Given a loss function $\rho_0^\star$ that satisfies as-*

sumption (A1) with $E_\Phi\{\rho_0^\star(u)\} = \bar{\epsilon}$*, set*

$$\rho^\star(t) = \rho_k(t) - \frac{E_\Phi\left\{2\rho_k(u) - \psi_k(u)u\right\}}{2\bar{\epsilon}}\rho_0^\star(t). \tag{27}$$

If the resulting $\rho^\star(t)$ *is nondecreasing on* $[0,\infty)$*, then* $\rho^\star$ *also satisfies assumption (A1) and the pair* $(\rho_0^*, \rho^\star)$ *is a solution to* (24)-(25)*.*

Theorem 3 allows us to find optimal pairs of loss functions within the class of extended one-step M-scale estimators. Note that in contrast to the class of M-scales, it follows from Theorem 3 that there is no unique solution within the class of extended one-step M-scale estimators anymore. We now investigate and compare possible solutions within this class.

Due to assumption (A1), the condition that $\rho^\star$ is nondecreasing on $[0,\infty)$, required in Theorem 3, can be replaced by the condition $\psi^\star(t) \geq 0$ on $[0,\infty)$ which in turn is equivalent to the condition

$$\frac{\psi_0^\star(t)}{2\bar{\epsilon}} \leq \frac{\psi_k(t)}{E_\Phi\left\{2\rho_k(u) - \psi_k(u)u\right\}}. \tag{28}$$

Once a pair $(\bar{\gamma}, \bar{\epsilon})$ has been chosen (with $\bar{\gamma} \in [1.167, \infty)$ and $\bar{\epsilon} \in [0, 0.5]$), it follows from Theorem 3 that for any initial loss function $\rho_0^\star$ such that (28) is satisfied, there exists a corresponding loss function $\rho^\star$, given by (27), such that the resulting extended one-step M-estimator of scale has maximal Gaussian efficiency among extended one-step M-scales satisfying the given bounds on GES and BDP. Note that condition (28) implies that the function $\rho_0^\star$ must be chosen so that the corresponding $\psi_0^\star$ becomes zero before ψ_k does. This, for example, excludes the use of Welsh and Cauchy loss functions (see Table 1).

As an illustration, we take the GES bound $\bar{\gamma} = 2.691$ corresponding to a maximal M-scale efficiency of 95% [$k(2.691) = 4.683$ in (26)]. Using (27), we can now construct extended one-step M-scales with 50% BDP and the same GES=2.691 that also reach the maximal efficiency of 95%. Figure 3 shows the optimal loss function $\rho^\star$ corresponding to an initial loss function $\rho_0^\star$ taken from the Huber, Yohai-Zamar, Tukey-biweight, and truncated linear families (see Table 1). Note that the corresponding optimal loss functions $\rho^\star$ look all very similar in the four considered cases and resemble well the optimal M-scale loss function ρ_k which is the dotted curve in these plots.

It follows from Theorem 3 that for several choices $\rho_0^\star$ that guarantee a desired breakdown point a corresponding $\rho^\star$ can be obtained such that the resulting extended one-step M-scales satisfy the desired bound on the GES and have maximal efficiency. According to (9), the influence function of extended one-step M-scales does not depend on the choice of the Fisher-consistent regression functionals. Hence, the result of Theorem 3 applies in particular to τ and MM-scales. Different optimal pairs of loss functions $(\rho_0^\star, \rho^\star)$ then lead to τ and MM-scales with the same GES and BDP. However, their maxbias behavior may be distinct. Therefore, we compare the maxbias behavior of τ and MM-scales based on different optimal pairs of loss functions. We consider bias curves

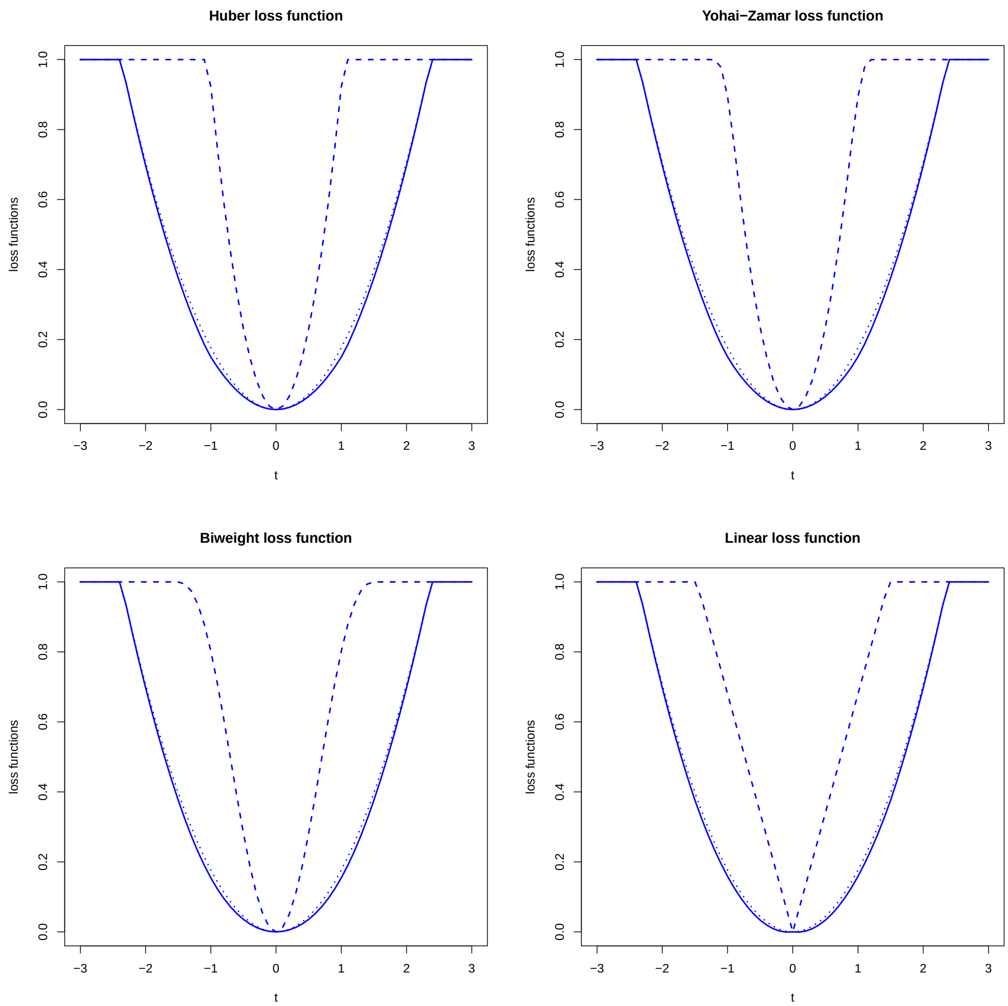

Figure 3: Pairs of loss functions ρ_0^* (dashed) and ρ^* (solid) that maximize the Gaussian efficiency (95%) of the extended one-step M-scale estimator subject to a bound on the GES ($\bar{\gamma} = 2.691$) and with maximal breakdown point ($\bar{\epsilon} = 50\%$). The loss function ρ_0^* is taken from the Huber, Yohai-Zamar, Tukey-biweight, and truncated linear families, respectively. The corresponding optimal M-scale loss function is shown in the plots as well (dotted line).

based on (20)-(21) corresponding to Corollary 3. However, note that $\rho^\star$ does not necessarily satisfy assumption (A3). In that case the bias expressions in (20)-(21) yield a lower bound for the maxbias which is not guaranteed to be tight.

Figure 4 shows the bias curves of the τ/MM-scales with 50% BDP, 95% efficiency and GES bound $\bar{\gamma} = 2.691$ as considered before. We only show the maximum generalized bias curves which are identical to the explosion maxbias curves in this case. The implosion maxbias curves are lower and virtually the same for the different optimal pairs of loss functions. From the plot it can be seen that the choice of ρ_0^* indeed affects the (max)bias behavior for larger contamination fractions. Not surprisingly, initial loss functions that yield S-scales with low maxbias result in Hampel-like optimal τ/MM-scales with lower maxbias as well.

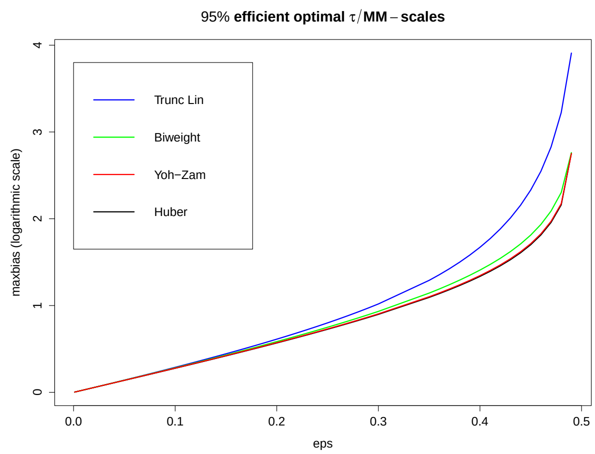

Figure 4: Maxbias curves for 50% BDP and 95% efficient τ/MM-scale functionals with GES bound $\bar{\gamma} = 2.691$. The legend of the plot lists the curves from top left to bottom right.

Note that the location-scale model is a special case of the regression model. Hence, our results also hold for this case. For instance, it follows from Theorem 3 that τ and MM-scale estimators can improve on Hampel's optimal M-scales in the sense that they can have the same (maximal) efficiency as Hampel's optimal M-scales subject to the bound on the GES, but at the same time they have a desired minimal breakdown point.

5. Regression and scale estimation

In the previous sections we have studied the robustness of residual scale estimators and found that fortunately the most robust and efficient scale estimators are those immediately derived from commonly used robust regression estimators: MM and τ-estimators. In practice, the loss functions for the MM and τ estimators are often tuned to reach a desired Gaussian efficiency for the regression parameter estimators. This tuning then automatically also determines the robustness and efficiency of the corresponding residual scale estimator. Therefore, in this section we compare the maxbias/efficiency performance of regression and scale MM and τ-estimators when the loss functions have been tuned to obtain 95% efficiency for the regression estimators.

In the case of MM and τ-estimators it turns out that the efficiency of the scale estimators is considerably larger than that of the associate regression estimators (over 98% for 95% efficient regression estimators), so we focus on the maxbias behavior of these estimators. For τ-estimators, expressions for the maxbias of the regression estimators have been derived by Berrendero and Zamar (2001). In the case of MM-estimators, upper and lower bounds for the maxbias of the regression estimators have been derived by Berrendero et al. (2007). Berrendero et al. (2007) also considered constrained M-estimators of regression (Mendes and Tyler, 1996) and have shown that these regression estimators have a very good maxbias behavior. It can thus be expected that the maxbias of their associated scale estimator is low as well, but has not been investigated so far. Since this scale estimator does not belong to our class of generalized M-scales, we do not include the constrained M-estimators in our comparison which is limited to robust regression estimators with a generalized M-scale as associated scale estimator.

Note that while the maxbias of MM and τ-scales in the linear model (1) does not depend on G_0, the distribution of the predictors, this is not the case anymore for the maxbias of the regression MM and τ-estimators. We consider the maxbias of the estimators when all the covariates and the error term follow normal distributions. Figure 5 compares the maxbias for the regression coefficients (left plots) and maxbias for the corresponding scale estimators (right plots) when both the regression MM and τ-estimators have been tuned to achieve 95% efficiency. For this comparison, we consider the Tukey biweight (top) and the Yohai-Zamar optimal (bottom) loss functions. Results for the Huber's truncated square loss function (not shown here) are similar to those for the optimal Yohai-Zamar function.

Note that the scale and regression maxbiases are always smaller when using the Yohai-Zamar optimal loss function. Hence, we restrict attention to this loss function. For this loss function, the maxbias for the scale τ-estimator is equal to or lower than the maxbias for the corresponding scale MM-estimator, except for very large fractions of contamination. Moreover, the maxbias for the regression τ-estimator is uniformly smaller than the maxbias for the regression MM-estimator. Therefore, taking efficiency as well as regression/scale maxbias into account, we conclude that τ estimators based on Yohai-Zamar loss functions can be recommended.

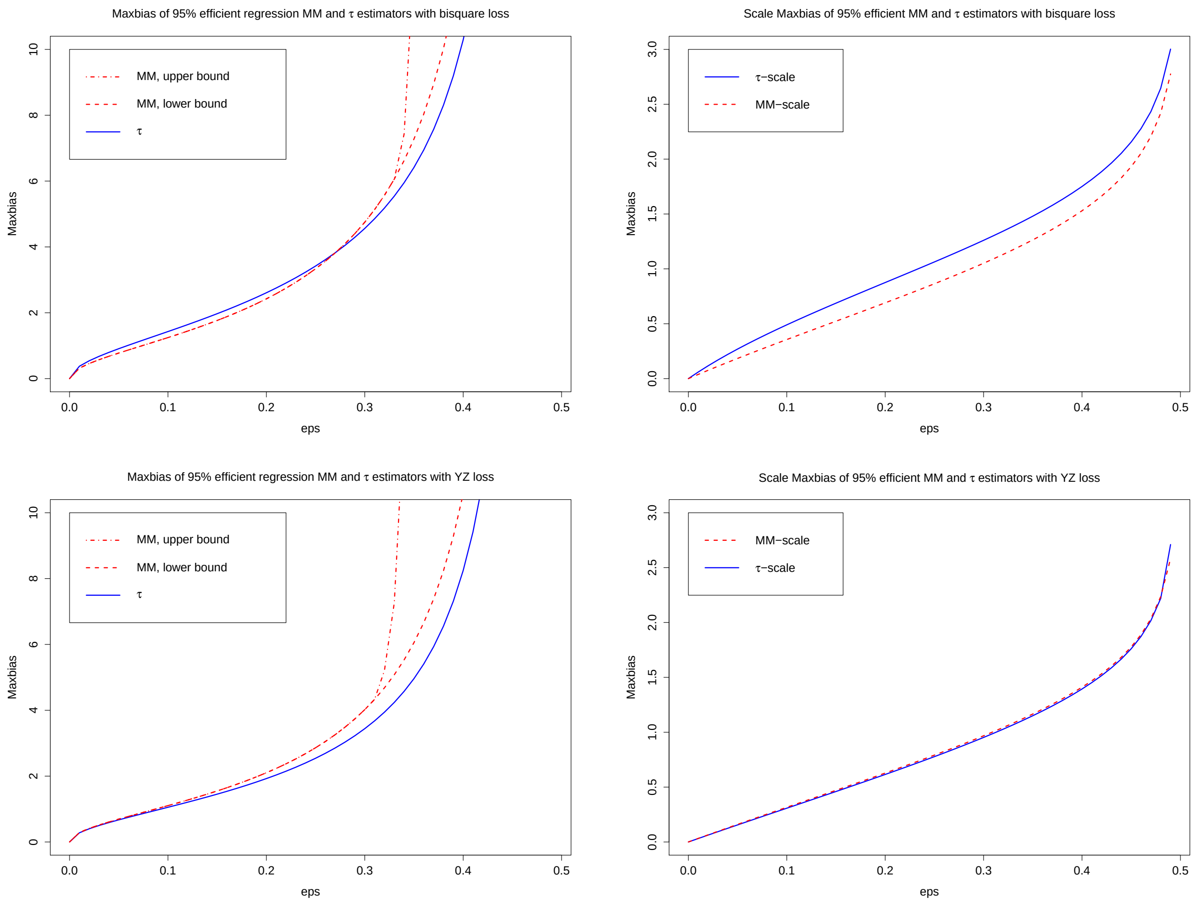

Figure 5: Maxbias comparisons for τ/MM regression (left) and scale (right) estimators with 95% regression efficiency. The top panel shows results for the Tukey biweight loss function, the bottom panel shows results for the Yohai-Zamar optimal loss function.

6. Appendix

Proof of Theorem 1. Using the notation $H_\epsilon = H_\epsilon^{(y_0,\mathbf{z}_0)}$ for simplicity, the influence function of $\sigma(H;\boldsymbol{\beta}(H))$ satisfies

$$
\begin{aligned}
IF(y_0,\mathbf{z}_0;\hat\sigma,H_0) &= \frac{\partial}{\partial\epsilon}\sigma(H_\epsilon;\hat s,\hat{\boldsymbol{\beta}})\bigg|_{\epsilon=0} = \frac{\partial}{\partial\epsilon}\sqrt{\sigma(H_\epsilon;\hat s,\hat{\boldsymbol{\beta}})^2}\,\bigg|_{\epsilon=0} \\
&= \frac{1}{2\sigma(H_0;\hat s,\hat{\boldsymbol{\beta}})}\,\frac{\partial}{\partial\epsilon}\sigma(H_\epsilon;\hat s,\hat{\boldsymbol{\beta}})^2\,\bigg|_{\epsilon=0} = \frac{1}{2}\,\frac{\partial}{\partial\epsilon}\sigma(H_\epsilon;\hat s,\hat{\boldsymbol{\beta}})^2\,\bigg|_{\epsilon=0}.
\end{aligned}
\tag{29}
$$

Using the assumptions in Theorem 1, we obtain from (3) that

$$
\begin{aligned}
\frac{\partial}{\partial\epsilon}\sigma(H_\epsilon;\hat s,\hat{\boldsymbol{\beta}})^2\,\bigg|_{\epsilon=0} &= \frac{\partial}{\partial\epsilon}\left\{\frac{\hat s^2(H_\epsilon)}{b}\left[(1-\epsilon)E_{H_0}\left\{\rho\left(\frac{r(\hat{\boldsymbol{\beta}}(H_\epsilon))}{\hat s(H_\epsilon)}\right)\right\}+\epsilon\rho\left(\frac{y_0-\mathbf{x}_0^t\hat{\boldsymbol{\beta}}(H_\epsilon)}{\hat s(H_\epsilon)}\right)\right]\right\}\Bigg|_{\epsilon=0} \\
&= 2\hat s(H_0)IF(y_0,\mathbf{z}_0;\hat s,H_0)+ \\
&\frac{\hat s^2(H_0)}{b}\,\frac{\partial}{\partial\epsilon}\left[(1-\epsilon)E_{H_0}\left\{\rho\left(\frac{r(\hat{\boldsymbol{\beta}}(H_\epsilon))}{\hat s(H_\epsilon)}\right)\right\}+\epsilon\rho\left(\frac{y_0-\mathbf{x}_0^t\hat{\boldsymbol{\beta}}(H_\epsilon)}{\hat s(H_\epsilon)}\right)\right]\Bigg|_{\epsilon=0} \\
&= 2\,IF(y_0,\mathbf{z}_0;\hat s,H_0)+\frac{1}{b}\left[-b+E_{H_0}\left(\psi(y)\,\frac{\partial}{\partial\epsilon}\frac{y-\mathbf{x}^t\hat{\boldsymbol{\beta}}(H_\epsilon)}{\hat s(H_\epsilon)}\bigg|_{\epsilon=0}\right)+\rho(y_0)\right] \\
&= 2\,IF(y_0,\mathbf{z}_0;\hat s,H_0)+\frac{\rho(y_0)-b}{b}+ \\
&\frac{1}{b}E_{H_0}\left[\psi(y)(-\mathbf{x}^tIF(y_0,\mathbf{z}_0;\hat{\boldsymbol{\beta}},H_0)-yIF(y_0,\mathbf{z}_0;\hat s,H_0))\right] \\
&= 2\,IF(y_0,\mathbf{z}_0;\hat s,H_0)+\frac{\rho(y_0)-b}{b}-\frac{E_{H_0}\{\psi(y)y\}}{b}IF(y_0,\mathbf{z}_0;\hat s,H_0)
\end{aligned}
$$

Inserting this expression in (29) proofs the theorem.

Proof of Theorem 2. We first derive the result for the explosion maxbias. Due to the symmetry of F_0, for any $s>0$ and $\mathbf{v}\in\mathbb{R}^p$ it holds that

$$
E_{F_0}\rho\left(\frac{y-\mathbf{x}^t\mathbf{v}}{s}\right)\geq E_{F_0}\rho\left(\frac{y}{s}\right). \tag{30}
$$

By (30), combining (A4) (ii) with (A3), and using that $\sup_t \rho(t) = \rho(\infty) = 1$, we obtain for any $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$ that

$$\begin{aligned}
g_{GM}(H_\epsilon^\infty; \mathbf{v}_1, \mathbf{v}_2) &= \frac{\hat{s}(H_\epsilon^\infty; \mathbf{v}_1)}{\sqrt{b}} \sqrt{(1-\epsilon) E_{H_0} \rho\left(\frac{y - \mathbf{x}^t \mathbf{v}_2}{\hat{s}(H_\epsilon^\infty; \mathbf{v}_1)}\right) + \epsilon} \\
&\geq \frac{\hat{s}(H_\epsilon^\infty; \mathbf{v}_1)}{\sqrt{b}} \sqrt{(1-\epsilon) E_{H_0} \rho\left(\frac{y}{\hat{s}(H_\epsilon^\infty; \mathbf{v}_1)}\right) + \epsilon} \\
&\geq \frac{\hat{s}(H_\epsilon^\infty; \mathbf{0})}{\sqrt{b}} \sqrt{(1-\epsilon) E_{H_0} \rho\left(\frac{y}{\hat{s}(H_\epsilon^\infty; \mathbf{0})}\right) + \epsilon} \\
&= g_{GM}(H_\epsilon^\infty; \mathbf{0}, \mathbf{0}) \qquad (31)
\end{aligned}$$

For any regression functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$ it follows from (31) that

$$\begin{aligned}
\sup_{H_\epsilon \in \mathcal{H}_\epsilon} \hat{\sigma}\left(H_\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}\right) &\geq \hat{\sigma}\left(H_\epsilon^\infty; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}\right) = g_{GM}\left(H_\epsilon^\infty, \tilde{\boldsymbol{\beta}}(H_\epsilon^\infty), \hat{\boldsymbol{\beta}}(H_\epsilon^\infty)\right) \\
&\geq g_{GM}\left(H_\epsilon^\infty; \mathbf{0}, \mathbf{0}\right) = \hat{\sigma}(H_\epsilon^\infty; \hat{s}(\mathbf{0}), \mathbf{0}),
\end{aligned}$$

which implies that

$$\sigma^+(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}) \geq \hat{\sigma}(H_\epsilon^\infty; \hat{s}(\mathbf{0}), \mathbf{0}). \qquad (32)$$

On the other hand, by combining (A4) (ii) with (A3), we obtain for any $H_\epsilon \in \mathcal{H}_\epsilon$ that

$$\begin{aligned}
\hat{\sigma}(H_\epsilon; \hat{s}(\mathbf{0}), \mathbf{0}) = g_{GM}(H_\epsilon; \mathbf{0}, \mathbf{0}) &= \frac{\hat{s}(H_\epsilon; \mathbf{0})}{\sqrt{b}} \sqrt{E_{H_\epsilon} \rho\left(\frac{y}{\hat{s}(H_\epsilon; \mathbf{0})}\right)} \\
&\leq \sqrt{\frac{1-\epsilon}{b} E_{H_0}\left[\hat{s}^2(H_\epsilon; \mathbf{0}) \rho\left(\frac{y}{\hat{s}(H_\epsilon; \mathbf{0})}\right)\right] + \frac{\epsilon}{b} \hat{s}^2(H_\epsilon; \mathbf{0})} \\
&\leq \sqrt{\frac{1-\epsilon}{b} E_{H_0}\left[\hat{s}^2(H_\epsilon^\infty; \mathbf{0}) \rho\left(\frac{y}{\hat{s}(H_\epsilon^\infty; \mathbf{0})}\right)\right] + \frac{\epsilon}{b} \hat{s}^2(H_\epsilon^\infty; \mathbf{0})} \\
&= \frac{\hat{s}(H_\epsilon^\infty; \mathbf{0})}{\sqrt{b}} \sqrt{E_{H_\epsilon^\infty} \rho\left(\frac{y}{\hat{s}(H_\epsilon^\infty; \mathbf{0})}\right)} \\
&= g_{GM}(H_\epsilon^\infty; \mathbf{0}, \mathbf{0}) = \hat{\sigma}(H_\epsilon^\infty; \hat{s}(\mathbf{0}), \mathbf{0}) \qquad (33)
\end{aligned}$$

Inequality (33) implies that

$$\begin{aligned}
\sigma^+(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}^{\text{opt}}), \hat{\boldsymbol{\beta}}^{\text{opt}}) = \sup_{H_\epsilon \in \mathcal{H}_\epsilon} \hat{\sigma}(H_\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}^{\text{opt}}), \hat{\boldsymbol{\beta}}^{\text{opt}}) &\leq \sup_{H_\epsilon \in \mathcal{H}_\epsilon} \hat{\sigma}(H_\epsilon; \hat{s}(\mathbf{0}), \mathbf{0}) \\
&\leq \hat{\sigma}(H_\epsilon^\infty; \hat{s}(\mathbf{0}), \mathbf{0}) \qquad (34)
\end{aligned}$$

Combining (32) with (34) yields

$$\sigma^+(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}^{\text{opt}}), \hat{\boldsymbol{\beta}}^{\text{opt}}) = \hat{\sigma}(H_\epsilon^\infty; \hat{s}(\mathbf{0}), \mathbf{0}) \leq \sigma^+(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}),$$

for any regression functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$, which proofs the first part of the theorem.

We now derive the implosion maxbias. By combining (A4) (i) with (A3) and using that $\rho(t) \geq 0$ for all $t \in \mathbb{R}$, we obtain for any $H_\epsilon \in \mathcal{H}_\epsilon$ and any $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$ that

$$\begin{aligned}
g_{GM}(H_\epsilon; \mathbf{v}_1, \mathbf{v}_2) &= \frac{\hat{s}(H_\epsilon, \mathbf{v}_1)}{\sqrt{b}} \sqrt{E_{H_\epsilon}\left\{\rho\left(\frac{y - \mathbf{x}^t\mathbf{v}_2}{\hat{s}(H_\epsilon, \mathbf{v}_1)}\right)\right\}} \\
&\geq \hat{s}(H_\epsilon, \mathbf{v}_1)\sqrt{\frac{1-\epsilon}{b}\left\{E_{H_0}\rho\left(\frac{y - \mathbf{x}^t\mathbf{v}_2}{\hat{s}(H_\epsilon, \mathbf{v}_1)}\right)\right\}} \\
&\geq \sqrt{\frac{1-\epsilon}{b}E_{H_0}\left\{\hat{s}^2(H_\epsilon, \mathbf{v}_1)\rho\left(\frac{y}{\hat{s}(H_\epsilon, \mathbf{v}_1)}\right)\right\}} \\
&\geq \sqrt{\frac{1-\epsilon}{b}E_{H_0}\left\{\hat{s}^2(H_\epsilon^0, \mathbf{0})\rho\left(\frac{y}{\hat{s}(H_\epsilon^0, \mathbf{0})}\right)\right\}} \\
&= \frac{\hat{s}(H_\epsilon^0, \mathbf{0})}{\sqrt{b}}\sqrt{E_{H_\epsilon^0}\rho\left(\frac{y}{\hat{s}(H_\epsilon^0, \mathbf{0})}\right)} \\
&= g_{GM}(H_\epsilon^0; \mathbf{0}, \mathbf{0}) \qquad (35)
\end{aligned}$$

From (35) it follows that for any regression functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$ we have that

$$\begin{aligned}
\inf_{H_\epsilon \in \mathcal{H}_\epsilon} \hat{\sigma}(H_\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}) &\geq \inf_{H_\epsilon \in \mathcal{H}_\epsilon} \inf_{\mathbf{v}_1, \mathbf{v}_2} g_{GM}(H_\epsilon; \mathbf{v}_1, \mathbf{v}_2) \\
&\geq g_{GM}(H_\epsilon^0; \mathbf{0}, \mathbf{0}) = \hat{\sigma}(H_\epsilon^0; \hat{s}(\mathbf{0}), \mathbf{0}),
\end{aligned}$$

which implies that $\sigma^-(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}) \geq \hat{\sigma}(H_\epsilon^0; \hat{s}(\mathbf{0}), \mathbf{0})$. If the regression functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$ satisfy the condition $\tilde{\boldsymbol{\beta}}(H_\epsilon^0) = \hat{\boldsymbol{\beta}}(H_\epsilon^0) = \mathbf{0}$, then we also have that $\sigma^-(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}) \leq \hat{\sigma}(H_\epsilon^0; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}) = \hat{\sigma}(H_\epsilon^0; \hat{s}(\mathbf{0}), \mathbf{0})$, hence $\sigma^-(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}}) = \hat{\sigma}(H_\epsilon^0; \hat{s}(\mathbf{0}), \mathbf{0})$ in this case.

Note that from (35) it also immediately follows that

$$\inf_{\mathbf{v}_1, \mathbf{v}_2} g_{GM}(H_\epsilon^0; \mathbf{v}_1, \mathbf{v}_2) \geq g_{GM}(H_\epsilon^0; \mathbf{0}, \mathbf{0}),$$

which implies that $\tilde{\boldsymbol{\beta}}^{\text{opt}}(H_\epsilon^0) = \hat{\boldsymbol{\beta}}^{\text{opt}}(H_\epsilon^0) = \mathbf{0}$. Hence, we have that $\sigma^-(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}^{\text{opt}}), \hat{\boldsymbol{\beta}}^{\text{opt}}) = \hat{\sigma}(H_\epsilon^0; \hat{s}(\mathbf{0}), \mathbf{0}) \leq \sigma^-(\epsilon; \hat{s}(\tilde{\boldsymbol{\beta}}), \hat{\boldsymbol{\beta}})$ for any regression functionals $\tilde{\boldsymbol{\beta}}(H)$ and $\hat{\boldsymbol{\beta}}(H)$, and the latter inequality becomes an equality if the condition $\tilde{\boldsymbol{\beta}}(H_\epsilon^0) = \hat{\boldsymbol{\beta}}(H_\epsilon^0) = \mathbf{0}$ is satisfied. This proofs the second part of the theorem.

Least Trimmed Squares Scales

Take any $\mathbf{v} \in \mathbb{R}^p$ and distribution H and define $\sigma_A^2(H; \mathbf{v}) = \int_A (y - \mathbf{x}^t\mathbf{v})^2\, dH$ for any measurable and bounded set $A \subset \mathbb{R}^p$ with $P_H(A) = 1 - \alpha$ where $0 <$

$\alpha < 1$ is the chosen trimming fraction. We call a set $\hat{A}_{\mathbf{v}}$ an LTS-solution if $\sigma^2_{\hat{A}_{\mathbf{v}}}(H;\mathbf{v}) \leq \sigma^2_A(H;\mathbf{v})$ for all measurable and bounded sets A with $P_H(A) = 1-\alpha$. If H is discontinuous such that $P_H(A) \neq 1-\alpha$ for all measurable and bounded sets $A \subset \mathbb{R}^p$, then the above definition can easily be adjusted as in Croux and Haesbroeck (1999). The LTS-scale functional can now be defined by $\hat{\sigma}^2_{LTS(\alpha)}(H;\mathbf{v}) = c_\alpha \sigma^2_{\hat{A}_{\mathbf{v}}}(H;\mathbf{v})$ where c_α can be chosen to make the scale $\hat{\sigma}_{LTS(\alpha)}(H_0;\boldsymbol{\beta}_0)$ a consistent estimator of σ_0.

It can easily be seen that the set $\mathcal{E}(H;\mathbf{v}) = \{(y,\mathbf{z}) \in \mathbb{R}^p; (y-\mathbf{x}^t\mathbf{v})^2 \leq d^2_{\mathbf{v},\alpha}(H)\}$ where $d^2_{\mathbf{v},\alpha}(H)$ is chosen such that $P_H(\mathcal{E}(H;\mathbf{v})) = 1-\alpha$ is an LTS solution. Indeed, for any measurable and bounded set A with $P_H(A) = 1-\alpha$ we have that

$$\begin{aligned}
\sigma^2_A(H;\mathbf{v}) = \int_A (y-\mathbf{x}^t\mathbf{v})^2\,dH &= \int_{A\cap\mathcal{E}(H;\mathbf{v})} (y-\mathbf{x}^t\mathbf{v})^2\,dH + \int_{A\setminus\mathcal{E}(H;\mathbf{v})} (y-\mathbf{x}^t\mathbf{v})^2\,dH \\
&\geq \int_{A\cap\mathcal{E}(H;\mathbf{v})} (y-\mathbf{x}^t\mathbf{v})^2\,dH + d^2_{\mathbf{v},\alpha}(H) P_H(A \setminus \mathcal{E}(H;\mathbf{v})) \\
&= \int_{A\cap\mathcal{E}(H;\mathbf{v})} (y-\mathbf{x}^t\mathbf{v})^2\,dH + d^2_{\mathbf{v},\alpha}(H) P_H(\mathcal{E}(H;\mathbf{v}) \setminus A) \\
&\geq \int_{A\cap\mathcal{E}(H;\mathbf{v})} (y-\mathbf{x}^t\mathbf{v})^2\,dH + \int_{\mathcal{E}(H;\mathbf{v})\setminus A} (y-\mathbf{x}^t\mathbf{v})^2\,dH \\
&= \int_{\mathcal{E}(H;\mathbf{v})} (y-\mathbf{x}^t\mathbf{v})^2\,dH = \sigma^2_{\mathcal{E}(H;\mathbf{v})}(H;\mathbf{v}).
\end{aligned}$$

We thus obtain that $\hat{\sigma}^2_{LTS(\alpha)}(H;\mathbf{v}) = c_\alpha \sigma^2_{\mathcal{E}(H;\mathbf{v})}(H;\mathbf{v})$.

Note that the regression LTS functional $\hat{\boldsymbol{\beta}}_{LTS(\alpha)}$ can be defined as $\hat{\boldsymbol{\beta}}_{LTS(\alpha)}(H) = \operatorname{argmin}_{\mathbf{v}} \hat{\sigma}^2_{LTS(\alpha)}(H;\mathbf{v})$. It has been shown that $\hat{\boldsymbol{\beta}}_{LTS}$ has a breakdown point given by $\min(\alpha, 1-\alpha)$ and is Fisher-consistent at H_0, i.e. $\hat{\boldsymbol{\beta}}_{LTS(\alpha)}(H_0) = \mathbf{0}$ (see e.g. Agulló et al., 2008).

Lemma 1. *Suppose that assumption (A2) holds and to avoid degenerate situations, we assume that $P_{H_0}(y - \mathbf{x}^t\mathbf{v} = 0) < 1-\alpha$ for any $\mathbf{v} \neq \mathbf{0}$. Then, the LTS-scale functional $\hat{\sigma}_{LTS(\alpha)}(H;\mathbf{v})$ satisfies conditions (A4).*

Proof of Lemma 1. For any $\mathbf{v} \in \mathbb{R}^p$ and $H_\epsilon \in \mathcal{H}_\epsilon$ we have that

$$1-\alpha = P_{H_\epsilon}(\mathcal{E}(H_\epsilon;\mathbf{v})) \leq (1-\epsilon)P_{H_0}(\mathcal{E}(H_\epsilon;\mathbf{v})) + \epsilon$$

which implies that $P_{H_0}(\mathcal{E}(H_\epsilon;\mathbf{v})) \geq (1-\alpha-\epsilon)/(1-\epsilon)$. Let us define $q^-_{\mathbf{v},\epsilon}$ such that $P_{H_0}(\{(y-\mathbf{x}^t\mathbf{v})^2 \leq q^-_{\mathbf{v},\epsilon}\}) = (1-\alpha-\epsilon)/(1-\epsilon)$, then clearly $d^2_{\mathbf{v},\alpha}(H_\epsilon) \geq q^-_{\mathbf{v},\epsilon}$.

It follows that

$$\begin{aligned}\hat{\sigma}^2_{LTS(\alpha)}(H_\epsilon;\mathbf{v}) = c_\alpha \sigma^2_{\mathcal{E}(H_\epsilon;\mathbf{v})}(H_\epsilon;\mathbf{v}) &= c_\alpha \int_{\mathcal{E}(H_\epsilon;\mathbf{v})} (y-\mathbf{x}^t\mathbf{v})^2\, dH_\epsilon \\ &\geq c_\alpha(1-\epsilon)\int_{\mathcal{E}(H_\epsilon;\mathbf{v})} (y-\mathbf{x}^t\mathbf{v})^2\, dH_0 \\ &\geq c_\alpha(1-\epsilon)\int_{\{(y-\mathbf{x}^t\mathbf{v})^2\leq q^-_{\mathbf{v},\epsilon}\}} (y-\mathbf{x}^t\mathbf{v})^2\, dH_0 \end{aligned} \tag{36}$$

On the other hand, we have that

$$\hat{\sigma}^2_{LTS(\alpha)}(H^0_\epsilon;\mathbf{0}) = c_\alpha \sigma^2_{\mathcal{E}(H^0_\epsilon;\mathbf{0})}(H^0_\epsilon;\mathbf{0}) = c_\alpha(1-\epsilon)\int_{\mathcal{E}(H^0_\epsilon;\mathbf{0})} y^2\, dH_0, \tag{37}$$

where $1-\alpha = P_{H^0_\epsilon}(\mathcal{E}(H^0_\epsilon;\mathbf{0})) = (1-\epsilon)P_{H_0}(\mathcal{E}(H^0_\epsilon;\mathbf{0}))+\epsilon$. Hence, $P_{H_0}(\mathcal{E}(H^0_\epsilon;\mathbf{0})) = P_{H_0}(\{y^2 \leq d^2_{\mathbf{0},\alpha}(H^0_\epsilon)\}) = (1-\alpha-\epsilon)/(1-\epsilon)$, so that $d^2_{\mathbf{0},\alpha}(H^0_\epsilon) = q^-_{\mathbf{0},\epsilon}$ in this case.

Because the LTS regression functional is Fisher-consistent at H_0 for any $0<\alpha<1$, we have for $\alpha_\epsilon = \alpha/(1-\epsilon)$ that

$$\hat{\sigma}^2_{LTS(\alpha_\epsilon)}(H_0;\mathbf{0}) = c_{\alpha_\epsilon}\int_{\{y^2\leq q^-_{\mathbf{0},\epsilon}\}} y^2\, dH_0 \leq c_{\alpha_\epsilon}\int_{\{(y-\mathbf{x}^t\mathbf{v})^2\leq q^-_{\mathbf{v},\epsilon}\}} (y-\mathbf{x}^t\mathbf{v})^2\, dH_0,$$

for any $\mathbf{v}\in\mathbb{R}^p$. Combining this inequality with (36) and (37) yields $\hat{\sigma}_{LTS(\alpha)}(H_\epsilon;\mathbf{v}) \geq \hat{\sigma}_{LTS(\alpha)}(H^0_\epsilon;\mathbf{0})$ for any $\mathbf{v}\in\mathbb{R}^p$ and $H_\epsilon\in\mathcal{H}_\epsilon$, which shows part (i) of (A4).

To show (ii), first note that for any $H_\epsilon\in\mathcal{H}_\epsilon$ it holds that

$$1-\alpha = P_{H_\epsilon}(\mathcal{E}(H_\epsilon;\mathbf{0})) \geq (1-\epsilon)P_{H_0}(\mathcal{E}(H_\epsilon;\mathbf{0})),$$

which implies that $P_{H_0}(\mathcal{E}(H_\epsilon;\mathbf{0})) \leq (1-\alpha)/(1-\epsilon)$. Let us define $q^+_{\mathbf{v},\epsilon}$ such that $P_{H_0}(\{(y-\mathbf{x}^t\mathbf{v})^2 \leq q^+_{\mathbf{v},\epsilon}\}) = (1-\alpha)/(1-\epsilon)$, then clearly $d^2_{\mathbf{0},\alpha}(H_\epsilon)\leq q^+_{\mathbf{0},\epsilon}$. Therefore, $\hat{\sigma}^2_{LTS(\alpha)}(H_\epsilon;\mathbf{0})$ can be rewritten as

$$\begin{aligned}&\hat{\sigma}^2_{LTS(\alpha)}(H_\epsilon;\mathbf{0}) = c_\alpha\sigma^2_{\mathcal{E}(H_\epsilon;\mathbf{0})}(H_\epsilon;\mathbf{0}) = c_\alpha\left((1-\epsilon)\int_{\mathcal{E}(H^0_\epsilon;\mathbf{0})} y^2\, dH_0 + \epsilon\int_{\mathcal{E}(H^0_\epsilon;\mathbf{0})} y^2\, dH^*\right) \\ &= c_\alpha\left((1-\epsilon)\int_{\{y^2\leq q^+_{\mathbf{0},\epsilon}\}} y^2\, dH_0 - (1-\epsilon)\int_{\{d^2_{\mathbf{0},\alpha}(H_\epsilon)\leq y^2\leq q^+_{\mathbf{0},\epsilon}\}} y^2\, dH_0 + \epsilon\int_{\mathcal{E}(H^0_\epsilon;\mathbf{0})} y^2\, dH^*\right) \\ &= c_\alpha\left((1-\epsilon)\int_{\{y^2\leq q^+_{\mathbf{0},\epsilon}\}} y^2\, dH_0 - I_1 + I_2\right). \end{aligned} \tag{38}$$

For I_2 we obtain that

$$I_2 = \epsilon\int_{\mathcal{E}(H^0_\epsilon;\mathbf{0})} y^2\, dH^* \leq \epsilon\, d^2_{\mathbf{0},\alpha}(H_\epsilon)P_{H^*}(\mathcal{E}(H^0_\epsilon;\mathbf{0})) = \epsilon\, d^2_{\mathbf{0},\alpha}(H_\epsilon)P_{H^*}(\{y^2\leq d^2_{\mathbf{0},\alpha}(H_\epsilon)\}) \tag{39}$$

For I_1 we obtain that

$$\begin{aligned}
I_1 &= (1-\epsilon)\int_{\{d^2_{\mathbf{0},\alpha}(H_\epsilon)\le y^2\le q^+_{\mathbf{0},\epsilon}\}} y^2\, dH_0 \\
&\ge (1-\epsilon) d^2_{\mathbf{0},\alpha}(H_\epsilon) P_{H_0}(\{d^2_{\mathbf{0},\alpha}(H_\epsilon)\le y^2\le q^+_{\mathbf{0},\epsilon}\}) \\
&= (1-\epsilon) d^2_{\mathbf{0},\alpha}(H_\epsilon)\left(P_{H_0}(\{y^2\le q^+_{\mathbf{0},\epsilon}\}) - P_{H_0}(\{y^2\le d^2_{\mathbf{0},\alpha}(H_\epsilon)\})\right) \\
&= (1-\epsilon) d^2_{\mathbf{0},\alpha}(H_\epsilon)\left((1-\alpha)/(1-\epsilon) - P_{H_0}(\{y^2\le d^2_{\mathbf{0},\alpha}(H_\epsilon)\})\right) \\
&= d^2_{\mathbf{0},\alpha}(H_\epsilon)\left(1-\alpha-(1-\epsilon)P_{H_0}(\{y^2\le d^2_{\mathbf{0},\alpha}(H_\epsilon)\})\right) \\
&= \epsilon\, d^2_{\mathbf{0},\alpha}(H_\epsilon) P_{H^*}(\{y^2\le d^2_{\mathbf{0},\alpha}(H_\epsilon)\})
\end{aligned} \tag{40}$$

Using (39) and (40) in (38) yields

$$\hat\sigma^2_{LTS(\alpha)}(H_\epsilon;\mathbf{0}) \le c_\alpha (1-\epsilon)\int_{\{y^2\le q^+_{\mathbf{0},\epsilon}\}} y^2\, dH_0. \tag{41}$$

On the other hand, we have that

$$\hat\sigma^2_{LTS(\alpha)}(H_\epsilon^\infty;\mathbf{v}) = c_\alpha \sigma^2_{\mathcal{E}(H_\epsilon^\infty;\mathbf{v})}(H_\epsilon^\infty;\mathbf{v}) = c_\alpha(1-\epsilon)\int_{\mathcal{E}(H_\epsilon^\infty;\mathbf{v})} (y-\mathbf{x}^t\mathbf{v})^2\, dH_0, \tag{42}$$

where $1-\alpha = P_{H_\epsilon^\infty}(\mathcal{E}(H_\epsilon^\infty;\mathbf{v})) = (1-\epsilon)P_{H_0}(\mathcal{E}(H_\epsilon^\infty;\mathbf{v})$. Hence, $P_{H_0}(\mathcal{E}(H_\epsilon^\infty;\mathbf{v})) = P_{H_0}(\{(y-\mathbf{x}^t\mathbf{v})^2\le d^2_{\mathbf{v},\alpha}(H_\epsilon^\infty)\}) = (1-\alpha)/(1-\epsilon)$, so that $d^2_{\mathbf{v},\alpha}(H_\epsilon^\infty) = q^+_{\mathbf{v},\epsilon}$ in this case.

Fisher-consistency of the LTS regression functional at H_0 implies that for $\alpha_\epsilon = (\alpha-\epsilon)/(1-\epsilon)$ it holds that

$$\hat\sigma^2_{LTS(\alpha_\epsilon)}(H_0;\mathbf{0}) = c_{\alpha_\epsilon}\int_{\{y^2\le q^+_{\mathbf{0},\epsilon}\}} y^2\, dH_0 \le c_{\alpha_\epsilon}\int_{\{(y-\mathbf{x}^t\mathbf{v})^2\le q^+_{\mathbf{v},\epsilon}\}} (y-\mathbf{x}^t\mathbf{v})^2\, dH_0,$$

for any $\mathbf{v}\in\mathbb{R}^p$. Combining this inequality with (41) and (42) yields $\hat\sigma_{LTS(\alpha)}(H_\epsilon;\mathbf{0}) \le \hat\sigma_{LTS(\alpha)}(H_\epsilon^\infty;\mathbf{v})$ for any $\mathbf{v}\in\mathbb{R}^p$ and $H_\epsilon\in\mathcal{H}_\epsilon$, which completes the proof.

Lemma 2. *Suppose that assumptions (A1)-(A3) hold. Then, for any $\mathbf{v}\in\mathbb{R}^p$ and $H\in\mathcal{H}_\epsilon$ the M-scale functional $\hat\sigma^M(H;\mathbf{v})$ corresponding to (4) which is defined as*

$$E_H\left\{\rho\left(\frac{y-\mathbf{x}^t\mathbf{v}}{\hat\sigma^M(H;\mathbf{v})}\right)\right\} = b. \tag{43}$$

satisfies conditions (A4).

Proof of Lemma 2. Definition (43) implies that for any $\mathbf{v}\in\mathbb{R}^p$ and $H_\epsilon\in\mathcal{H}_\epsilon$ we have that

$$b = E_{H_\epsilon}\left\{\rho\left(\frac{y-\mathbf{x}^t\mathbf{v}}{\hat\sigma^M(H_\epsilon;\mathbf{v})}\right)\right\} \ge (1-\epsilon)E_{H_0}\left\{\rho\left(\frac{y-\mathbf{x}^t\mathbf{v}}{\hat\sigma^M(H_\epsilon;\mathbf{v})}\right)\right\}$$

$$\geq (1-\epsilon) E_{H_0}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon;\mathbf{v})}\right)\right\} \tag{44}$$

Similarly, for $\hat{\sigma}_M(H_\epsilon^0;\mathbf{0})$ we have that

$$b = E_{H_\epsilon^0}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon^0;\mathbf{0})}\right)\right\} = (1-\epsilon) E_{H_0}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon^0;\mathbf{0})}\right)\right\} \tag{45}$$

From (44) and (45) we obtain that

$$E_{H_0}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon;\mathbf{v})}\right)\right\} \leq E_{H_0}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon^0;\mathbf{0})}\right)\right\}.$$

Together with (A1) this inequality implies that $\hat{\sigma}_M(H_\epsilon;\mathbf{v}) \geq \hat{\sigma}_M(H_\epsilon^0;\mathbf{0})$ for any $\mathbf{v}\in\mathbb{R}^p$ and $H_\epsilon \in \mathcal{H}_\epsilon$, which shows the first inequality in (i).

To show (ii), note that for any $H_\epsilon \in \mathcal{H}_\epsilon$, $\sigma_M(H_\epsilon;\mathbf{0})$ satisfies

$$b = E_{H_\epsilon}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon;\mathbf{0})}\right)\right\} \leq (1-\epsilon) E_{H_0}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon;\mathbf{0})}\right)\right\} + \epsilon. \tag{46}$$

On the other hand, since $\sup\rho = \rho(\infty) = 1$, we have that $\hat{\sigma}_M(H_\epsilon^\infty;\mathbf{v})$ satisfies

$$b = (1-\epsilon) E_{H_0}\left\{\rho\left(\frac{y-\mathbf{x}^t\mathbf{v}}{\hat{\sigma}^M(H_\epsilon^\infty;\mathbf{v})}\right)\right\} + \epsilon \geq (1-\epsilon) E_{H_0}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon^\infty;\mathbf{v})}\right)\right\} + \epsilon \tag{47}$$

Combining (46) with (47) yields

$$E_{H_0}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon;\mathbf{0})}\right)\right\} \geq E_{H_0}\left\{\rho\left(\frac{y}{\hat{\sigma}^M(H_\epsilon^\infty;\mathbf{v})}\right)\right\}$$

which implies that $\hat{\sigma}^M(H_\epsilon;\mathbf{0}) \leq \hat{\sigma}^M(H_\epsilon^\infty;\mathbf{v})$ for all $H_\epsilon \in \mathcal{H}_\epsilon$ and $\mathbf{v}\in\mathbb{R}^p$. This completes the proof.

Proof of Theorem 3. As explained in Section 4, we need to define the function ρ^* such that $\rho_k(t) = W_\Phi\, \rho_0^*(t) + \rho^*(t)$ with W_Φ given by (10). Therefore, $\rho^*(t)$ should satisfy the following equation

$$\rho^*(t) = \rho_k(t) - \frac{2E_\Phi\left\{\rho^*(u)\right\} - E_\Phi\left\{\psi^*(u)u\right\}}{E_\Phi\left\{\psi_0^*(u)u\right\}}\rho_0^*(t) \tag{48}$$

To eliminate ρ^* from the right hand side, we take derivatives on both sides of the above equation, then multiply by t and take expectations. This yields

$$E_\Phi\left\{\psi_k(u)u\right\} = 2\, E_\Phi\left\{\rho^*(u)\right\}. \tag{49}$$

On the other hand, just taking expectations in (48) gives

$$E_\Phi\left\{\rho^*(u)\right\} = E_\Phi\left\{\rho_k(u)\right\} - \frac{2E_\Phi\left\{\rho^*(u)\right\} - E_\Phi\left\{\psi^*(u)u\right\}}{E_\Phi\left\{\psi_0^*(u)u\right\}}\bar{\epsilon} \tag{50}$$

Using (49), we can rewrite (50) as

$$\frac{2E_\Phi[\{\rho^*(u)\} - E_\Phi\{\psi^*(u)u\}}{E_\Phi\{\psi_0^*(u)u\}} = \frac{E_\Phi\{\rho_k(u)\} - \frac{1}{2}E_\Phi\{\psi_k(u)u\}}{\bar{\epsilon}}.$$

By inserting this result into (48) we obtain that ρ^* is given by

$$\rho^*(t) = \rho_k(t) - \frac{2E_\Phi\{\rho_k(u)\} - E_\Phi\{\psi_k(u)u\}}{2\bar{\epsilon}}\rho_0^*(t)$$

which corresponds to (27).

References

Agulló, J., Croux, C., Van Aelst, S. (2008). The multivariate least-trimmed squares estimator. *Journal of Multivariate Analysis,* 99, 311–338.

Berrendero, J. R., Mendes, B. V. M., Tyler, D. E. (2007). On the maximum bias functions of MM-estimates and constrained M-estimates of regression. *Annals of Statistics,* 35, 13–40.

Berrendero, J. R., Zamar, R. H. (2001). Maximum bias curves for robust regression with non-elliptical regressors. *Annals of Statistics,* 29, 224–251.

Croux, C. (1994). Efficient High-Breakdown M-Estimators of Scale. *Statistics and Probability Letters*, 19, 371–379.

Croux, C., Haesbroeck, G. (1999). Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis,* 71, 161–190.

Croux, C., and Haesbroeck, G. (2001). Maxbias curves of robust scale estimators based on subranges. *Metrika*, 53, 101–122.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. John Wiley and Sons, New York.

He, X. and Simpson, D.G. (1993). Lower bounds for contamination bias: globally minimax versus locally linear estimation. *Annals of Statistics,* 21, 314–337.

Heritier, S. and Ronchetti, E. (1994). Robust bounded-influence tests in general parametric models. *Journal of the American Statistical Association*, 89, 897–904.

Huber, P.J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35, 73–101.

Huber, P.J. (1981). *Robust Statistics*. Wiley, New York.

Iglewicz, B. (1982), Robust scale estimates, *Understanding Robust and Exploratory Data Analysis* (eds. D.C. Hoaglin, F. Mosteller and J.W. Tukey), 404–431, John Wiley and Sons, New York.

Martin, R. D., Yohai, V. J., Zamar, R. (1989). Min-max bias robust regression. *Annals of Statistics,* 17, 1608–1630.

Martin, R.D. and Zamar, R.H. (1989). Asymptotically min-max bias-robust M-estimates of scale for positive random variables. *Journal of the American Statistical Association,* 84, 494–501.

Martin, R.D. and Zamar, R.H. (1993). Bias robust estimates of scale. *Annals of Statistics,* 21, 99–1017.

Mendes, B.V.M. and Tyler, D.E. (1996). Constrained M-estimation for regression. *Robust Statistics, Data Analysis and Computer Intensive Methods* (ed. H. Rieder), Lecture Notes in Statist. 109, 299–320, Springer, New York.

Randal, J.A. (2008). A reinvestigation of robust scale estimation in finite samples. *Computational Statistics and Data Analysis*, 52, 5014-5021.

Rousseeuw, P.J. and Croux, C. (1993). Alternatives to the Median Absolute Deviation. *Journal of the American Statistical Association,* 88, 1273–1283.

Rousseeuw, P. J., Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* Wiley-Interscience, New York.

Rousseeuw, P. J. and Yohai, V. J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series* (J. Franke, W. Hardle and D. Martin, eds.). Lecture Notes in Statist., 26, 256–272. Berlin: Springer-Verlag.

Van Aelst, S. and Willems, G. (2005). Multivariate regression S-estimators for robust estimation and inference. *Statistica Sinica,* 15, 981–1001.

Van Aelst, S. and Willems, G. (2011). Robust and efficient one-way MANOVA tests. *Journal of the American Statistical Association,* 106, 706-718.

Yohai, V.J. (1987). High breakdown point and high efficiency robust estimates for regression. *Annals of Statistics,* 15, 642–656.

Yohai, V.J. and Zamar, R.H. (1988). High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association,* 83, 406–413.

Yohai, V.J. and Zamar, R.H. (1997). Optimal locally robust M-estimates of regression *Journal of Statistical Planning and Inference,* 64, 309–323